SCIENTIFIC COMPUTING WORLD

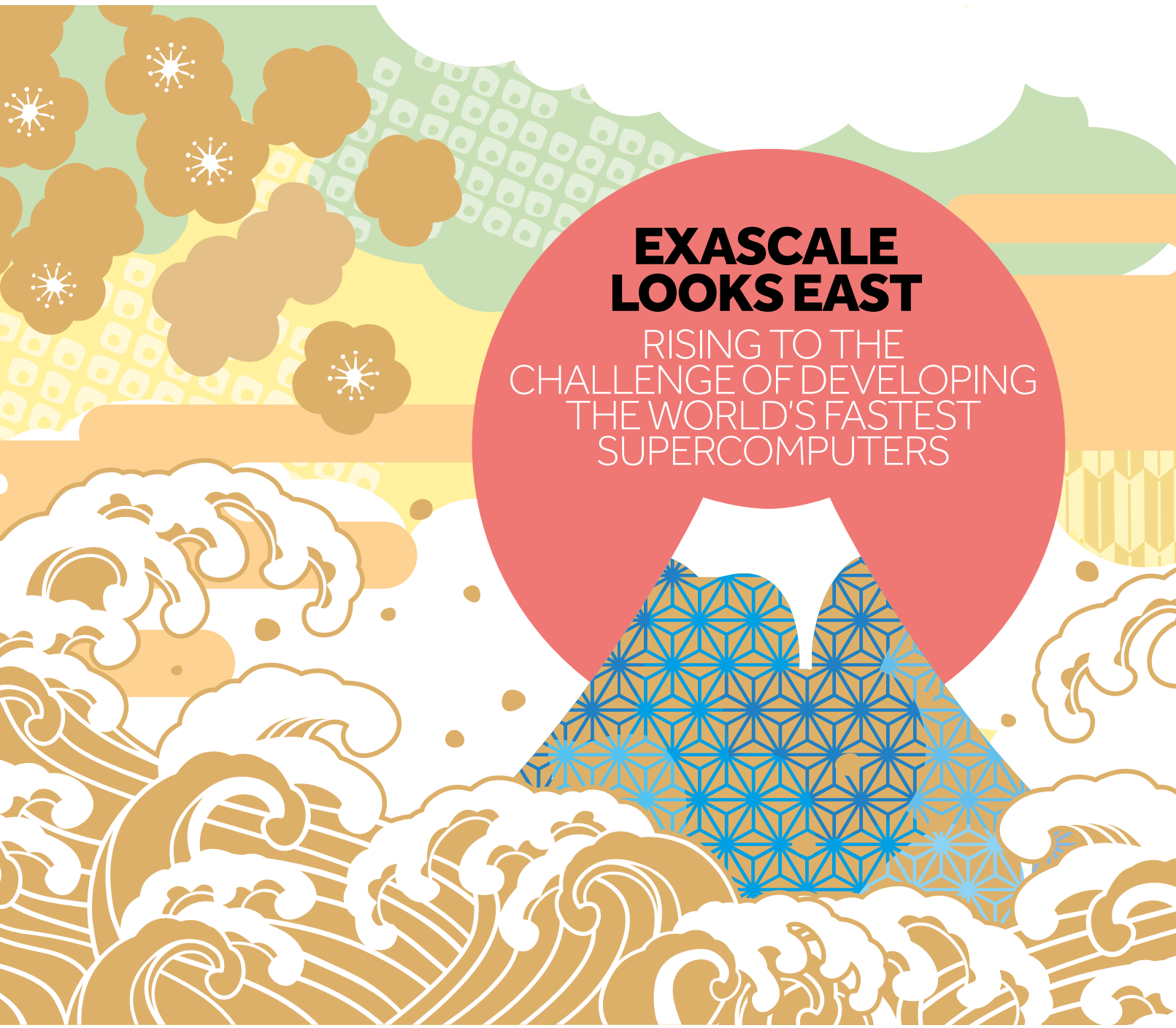
Computing solutions for scientists and engineers

Autumn 2020
Issue #173

High performance computing
Engineering demands

Laboratory informatics
Transform the laboratory

Modelling and simulation
Expanding CFD simulation



EXASCALE LOOKS EAST

RISING TO THE CHALLENGE OF DEVELOPING THE WORLD'S FASTEST SUPERCOMPUTERS

www.scientific-computing.com

Full Program. Fully Virtual.

Tutorials	Workshops	Program	Exhibits
9–10	11–13	16–19	17–19
NOVEMBER	NOVEMBER	NOVEMBER	NOVEMBER

The International Conference for High Performance Computing, Networking, Storage, and Analysis

SC20

Everywhere we are | more than hpc.

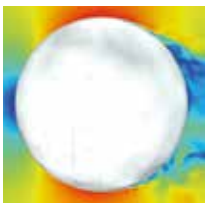
Register Today!

sc20.supercomputing.org

Sponsored by: IEEE COMPUTER SOCIETY | TCHPC | acm sighpc

Contents

Autumn 2020 • Issue 173



High performance computing

Laboratory informatics

Modelling and simulation

Resources

Editorial and administrative team
Editor: Robert Roe
robert.roe@europascience.com
Managing editor: Tim Gillett
editor.scw@europascience.com
Specialist reporters: Sophia Ktori, Gemma Church
Design: David Houghton, Zöe Andrews
Partnership & events executive: Charlie Mitchell
charlie.mitchell@europascience.com

Advertising team
Advertising manager: Mike Nelson
mike.nelson@europascience.com +44 (0) 1223 221039
Production manager: David Houghton
david.houghton@europascience.com +44 (0) 1223 221034

Corporate team
Managing director: Warren Clark

Scientific Computing World is published by Europa Science Ltd, 4 Signet Court, Cambridge, CB5 8LA • ISSN 1744-8026
Tel: +44 (0) 1223 211170 • Fax: +44 (0) 1223 213385
Web: www.researchinformation.info

Subscriptions: Free registration is available to qualifying individuals. Register online at www.scientific-computing.com Subscriptions £180 a year for six issues to readers outside registration requirements. Single issue £20. Orders to ESL, SCW Circulation, 4 Signet Court, Swann Road, Cambridge, CB5 8LA, UK. Tel: +44 (0)1223 221 030. Fax: +44 (0)1223 213 385. ©2020 Europa Science Ltd.

US copies: Scientific Computing World (ISSN 1356-7853/USPS No 018-753) is published bi-monthly for £180 per year by Europa Science Ltd, and distributed in the USA by DSW, 75 Aberdeen Rd, Emigsville PA 17318-0437. Periodicals postage paid at Emigsville PA. Postmaster: Send address corrections to Scientific Computing World PO Box 437, Emigsville, PA 17318-0437.

Cover: Shutterstock.com

Subscribe for free at
www.scientific-computing.com/subscribe

Robert Roe
Editor

Taking it up an exa-notch

In this issue we take a look at how data is driving the next generation of computing technologies. On page 4, we begin with exascale and the challenges facing researchers trying to reach this milestone across. With the first US-based exascale supercomputer delayed due to Intel's manufacturing issues, AMD will deploy the first system in the US.

Next, we have two technology focus on page 8 and 11 which detail software tools for HPC and cloud technologies. Following that on page 14 is an article which highlights the need for HPC technology in engineering simulation.

The laboratory informatics content starts on page 16 with a look at how big data is helping research into patients with sleep problems. On page 17, Sophia Ktori considers digital transformation in the laboratory, while the focus shifts to Covid-19 research on page 22, and how software tools can help expedite drug and vaccine development.

The importance of cloud computing and improving digital infrastructure is the focus of an article from Jisc's Matthew Dovey on page 25. Following that Synthace co-founder Markus Gershater discusses the role of Antha in developing a more effective platform for the design of experiments.

Maximising the use of laboratory data is the theme of an interview with Merck's Jean-Charles Wirth on page 27. The focus then shifts to biological data on page 28, with the role of software tools for data management under the spotlight.

On page 30 Gemma Church takes a deep dive into the increasing use of CFD in engineering.

@scwmagazine

FREE ON-DEMAND WEBCAST*
AVAILABLE NOW

Enabling Discovery In Research With Cloud Technologies

SPEAKERS

Prof. Imre Berger
Director, Max Planck Centre for Minimal Biology, University of Bristol, UK

Dr Dan Ruderman
Director of Analytics and Machine Learning, at Lawrence J. Ellison Institute for Transformative Medicine of USC, USA

Eric Grancher
Head of database services group, CERN IT department, at CERN, Switzerland

Taylor Newill
Senior Manager, HPC Product Management, at Oracle

Research drives the global expansion of knowledge and spurs scientific, technological and social innovation. Currently, all too often research investigations are constrained by the limitations of only using traditionally available computing technologies. Existing technologies may not be the most advanced, they may have capacity thresholds and typically will have a queuing system in place. These conditions impede progress and impact research outcomes. Cloud technologies, by contrast, are now opening up entirely new possibilities, enabling researchers to access additional, powerful resources and alternative technologies, exactly as and when they are required. Researchers today thus have many more options to help them accelerate time to discovery and advance innovation.

Listen to fascinating insights about scientists and research organisations using, or planning to use, cloud technologies, to advance their ground-breaking research endeavours.

SCIENTIFIC COMPUTING WORLD

ORACLE

www.scientific-computing.com/webcasts

*This webcast is FREE to view, though registration is required

LABWARE

LIMS and ELN in a single platform
Automation software for all laboratories

LABWARE
Results Count

Offices worldwide supporting customers
in more than 125 countries

www.labware.com

The challenges of exascale development

ROBERT ROE TAKES A LOOK AT CHALLENGES FACING THE FIRST GENERATION OF EXASCALE SUPERCOMPUTERS' DEVELOPMENT

With the recent announcement that the first scheduled US exascale system will be AMD-based due to the delays of the Intel-developed Aurora system, the challenges of developing these record-breaking HPC systems is causing setbacks in the original exascale timeline.

Intel has struggled to deliver the 7nm processors that will be used in the Aurora system, announcing delays earlier this year. It was thought that this may delay the Aurora system for some time but this has now been confirmed by an interview in *InsideHPC* with US Under Secretary for Science Paul Dabbar, commenting that 'They're getting very close to the first machines that are going to be delivered next year. And the first one is going to be at Oak Ridge.'

This statement makes it clear that it will be the Oak Ridge system Frontier which is delivered first, with the Aurora system at Argonne further delayed until Intel can overcome its manufacturing challenges to deliver the 7nm chips.

It is now understood that the Intel 7nm chips for PC users will shift to late 2022 or early 2023 while server-grade CPUs are now planned to launch in late 2023. The company's 7nm GPU codenamed Ponte Vecchio will also be delayed until 2023.

Chip rivals such as Nvidia and AMD are both delivering 7nm chips which have been manufactured by TSMC.

Dabbar noted: 'We are in discussions with Intel about that. I think we're feeling good about the overall machine. I can't go through exactly all the different options that we're looking at for the Argonne machine, but we have a good degree of confidence that not very long after the Oak Ridge machine – that will be delivered as part of a plan to have at least one (exascale) machine up in 2021 – but not long after that we will have the Aurora machine, the Argonne machine, also.

'The details are still being identified about exactly what we're going to go through with Intel and their microelectronics. But we have confidence that the machine will also be delivered, and will be delivered right behind Oak Ridge. Our major partners have different components of the hardware and software stack,' said Dabbar.

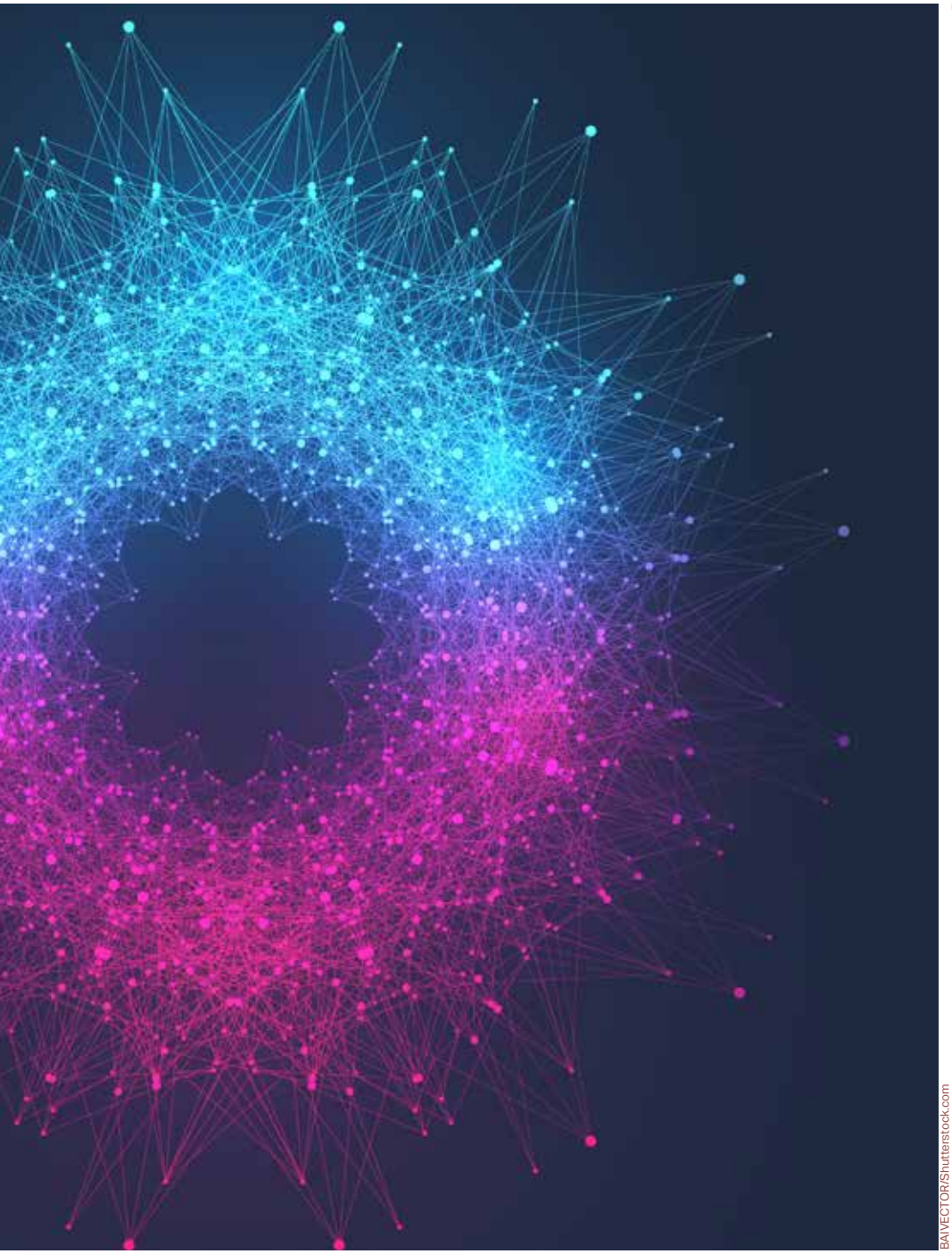
'One of the things that will be occurring with the exascale programme is the Shasta software stack that is … already developed to a large degree by HPE Cray. That's something they're developing as part of running their system. They actually have deployed early versions of that. Some of their other machines that they have that are not exascale, so they already have the earlier versions that have been de-risked.'

Dabbar stressed that a big part of the deployment process is focused on developing the integration between hardware and software. 'The software stack that is going to be riding on top of the hardware is going to be integrated. So a lot of the discussion today, and part of the deployment, is that layering of the operating system on top of the hardware.'

While it had been thought that the Intel-based system may be delayed since news about the 7nm chips was first announced earlier this year, this is the first confirmation that the exascale system developed by intel will be delayed. This highlights the difficulties in developing these systems that must deliver performance and reliability at an unprecedented scale.

Calculated risks

Professor Satoshi Matsuoka, director of Riken Center for Computational Science,



BAIVECTOR/Shutterstock.com

commented on the development of the Fugaku system and some of the challenges that the teams faced to deliver the hardware and software updates that made the system possible.

While it may not deliver the performance of the US systems scheduled to be delivered two years from now, the Fugaku supercomputer does highlight the development efforts of Japanese researchers and Fujitsu, who developed their own Arm-based processor and other technologies, such as the Tofu networking system used in Fugaku.

Matsuoka said early in development Riken saw the need to accommodate emerging technologies such as AI and maching learning (ML), which require mixed precision to accelerate performance. 'We have been very flexible in accommodating these types of emerging applications, and that is also due to the fact we are designing our own hardware from scratch.

'Designing any hardware is a balancing act, and then you try to leapfrog by using advanced technologies. In our case, that was another thing that we were able to do because we are a national project and we are trying to achieve this unobtainable goal.

'Otherwise, it would not be possible by the standard development processes of a single company,' he said.

Matsuoka gave standard CPU development cycles as an example. Noting that while a speedup of 10 to 20 per cent per generation is fine for a commercial company making incremental gains to support its existing customer base, it was not sufficient to realise the ambitions of the Fugaku supercomputer.

'It is harder to take risks, so you make advances but you make baby steps.' said Matsuoka.

'Of course, if you look at the holistic lifespan of hardware architecture you will see significant advances but that is 'still bound by the limitations of having to accommodate existing users'.

Commercial companies have to manage risk more carefully as the risk of new technology development falls solely to the organisation. In the case of Riken, the development of the processor was subsidised by the Japanese government and Fujitsu.

'In our case, we are a national project, we can, and in fact, we are obligated to take risks because the government are financing it. If we fail, then we may never be able to do it again but it is not like a company will go under. We are expected to take these risks and so we adopted or developed technologies that are otherwise still not used or utilised in existing CPUs such as high bandwidth memory (HBM). This has resulted in this leadership performance which others, to date, have not been able to follow,' added Matsuoka.

"They're getting very close to the first machines that are going to be delivered next year. The first one is going to be at Oak Ridge"

→

"We do not have the capability to develop a chip that has this level of reliability, but Fujitsu has been doing that for years"

Levchenko Ilia/Shutterstock.com

→ **Moonshot drives technology development**

While the development of exascale technologies are clearly much more challenging than standard HPC development, Matsuoka notes that the goals can be achieved through carefully planned partnerships.

'Commercial technologies have skills that we do not have. We cannot do detailed chip design. Fujitsu has been doing that for 50 to 60 years, if you count the mainframe days, so they have a wealth of experience, and the engineers and systems. Those are their assets, they are one of the few remaining companies left that can design server grade general-purpose CPUs.

'It was a great collaboration with Riken and other universities spearheading innovations, and these elements being realised as a tangible product. Unless you have a chip that is ultra-reliable then the machine will crash every second. If you leave it academia to make a chip it would be very likely that this would happen,' said Matsuoka.

'We do not have the capability to develop a chip that has this level of reliability but Fujitsu has been doing that for years. It was a fantastic combination of skillsets that allowed this moonshot-like development to come to fruition,' added Matsuoka.

The Riken centre has previously focused on the use of Sparc-based silicon for its previous supercomputers. The K computer, for example, the predecessor to the Fugaku supercomputer at Riken today, was based on Sparc processor technology developed at Fujitsu. While it may seem that switching from one architecture to another, which uses completely different Industry Standard Architecture (ISA), would suggest that much of the work would be replacing the old technology with new, there is significant technology transfer from the K computer to Fugaku.

'If you look at the processor architecture, what the processor has to do on a very basic level is to understand the instruction set. Instruction set architectures are like languages. You can convey the same meaning but there are different languages that you may choose to use,' said Matsuoka.

'However, if a language is only spoken by a thousand people, then you do not have the knowledge and resources that would be available in a more common language.'

'While Sparc, at one time, was very mainstream, in the heydays of Risk – maybe 20 to 30 years ago – the importance of Sparc diminished over time, and it has been replaced with X86 and other technologies,' noted Matsuoka. 'Fujitsu stuck with Sparc because they had considerable assets in Sparc-based codes.'

When planning the initial stages of Fugaku around ten years ago, Riken and its partners were looking to adopt a new ISA for the next generation of supercomputers.

'It would be a big risk and cost for Fujitsu to switch to another ISA, but, on the other hand, if they stuck to Sparc they would be diminished into obscurity,' stressed Matsuoka.

'For HPC it is critically important that we achieve generality, because we have gone mainstream.'

Matsuoka also noted that the ISA could have been x86 but 'there are some issues with x86 both at the system level and also from a licencing point of view, which made it almost impossible'.

He continued: 'Arm was the obvious choice because it had already become a standard in the embedded space. In the end, it was decided that we would use Arm and the rest is history,'

The legacy of the K computer

With this large change in technology and all the development that went into the new system and its associated technologies, what lessons or knowledge has been taken forward and helped the development of this new system?

From the perspective of professor Matsuoka, there has been a significant transfer of technology which benefits scientists, and also makes it easier for technology development in the future.

Matsuoka states: 'Then the question is, is there any legacy of K embedded into the machine or the architecture? The answer from the ISA perspective is, of course, no, because it is a completely different language, so to speak, but there are many other remnants that exist.

'The most pronounced is the fact that when Fujitsu designs its chips they divide it into the front and backend. The back end is fairly oblivious to the instruction set, because every language basically does the same thing. The backend of Fujitsu has been inherited over many years. It has been used for K, of course with evolutions, with gradual evolutions to add features but basically, it is the same backend.

'For Fugaku, the backend is an evolution of the design but there are commonalities with K's backend. It is not a direct re-utilisation, execution is much wider. There are memory enhancements and many other enhancements, but the baseline architecture is the same,' he said.

COMPUTING INSIGHT UK 2020

Science and Technology Facilities Council

Full programme will be issued soon - please check website for details

THURSDAY 3 DECEMBER 2020

www.stfc.ac.uk/ciuk

Registration is now open… and its FREE!

We are very excited to announce that **CIUK 2020** will be a completely online event - the first ever virtual CIUK conference—and registration is **FREE**! However, this does not mean that you will experience any less than at previous CIUK conferences. In fact by moving online we have an opportunity to include extra activities that we have not been able to include in any of the previous 30 editions of the event through an extended programme that will begin in November.

CIUK 2020 will include a full day of presentations, an online poster competition, and we will be presenting our Jacky Pallas Memorial Award to a young researcher who has completed an exciting and high impact research project.

We will also be launching the CIUK Student Cluster Challenge as planned – although we will move to an online format. Teams of six will compete a series of mini challenges over the months leading up to the final challenge on 3 December with a weekly league table and the eventual winners going forward to represent CIUK at the ISC Cluster Challenge in Germany next summer.

Why not get involved?

We have a number of calls currently open including the opportunity to submit an application to speak on the main programme and to submit a poster to our poster competition. We are also looking for nominations for the Jacky Pallas Memorial Award. If you would like to be more involved in CIUK 2020 all details can be found on our website… www.stfc.ac.uk/ciuk.

Tech focus: software tools

ZinetroN/Shutterstock.com

A ROUND-UP OF THE LATEST SOFTWARE TOOLS FOR SCIENTISTS AND RESEARCHERS USING HPC

As High-performance computing becomes a more ubiquitous tool for scientists and researchers in their daily lives, ensuring ease of use of these systems is important to accelerate productivity and scientific discovery.

Tools that help scientists manage their workloads and resources more effectively can help reduce the barrier to entry for new organisations or inexperienced users. Helping scientists push their workloads into the cloud, maximising the use of in-house resources or optimising the codes and increasing the efficiency are just some of the examples of benefits that help drive productivity and increase scientific output.

Debugging and code optimisation tools can help to reduce issues that may arise from software development. This has benefits in reducing the cost of code development and gives tools to scientific domain experts that help them better implement their scientific research to applications. This frees up time of expert staff and makes it easier for companies to pursue their computing research or product development, without investing as heavily in with coding and application development.

Making software tools accessible also helps to reduce the failure of workloads and helps management teams understand what has gone wrong when things fail. Whether it be hardware monitoring delivered in cluster management software or job scheduling services which report on workload performance, HPC teams can use this information to more accurately support users and deliver a more robust HPC service for scientists and engineers.

Commercial products

Adaptive Computing's Moab HPC Suite is a workload and resource orchestration platform that automates the scheduling, managing, monitoring and reporting of HPC workloads on a massive scale. Its patented intelligence engine uses multi-dimensional policies and advanced future modelling to optimise workload start and run times on diverse resources.

Nodus Cloud OS is intelligent cloud management technology that enables seamless access to compute resources, whether on-premise or in the cloud. Cloud deployment on any of the leading cloud providers becomes easily attainable with Nodus. It simplifies the complexities of running HPC and enterprise workloads and applications in the cloud.

Altair provides high-throughput computing solutions that help you orchestrate, visualise, optimise and analyse in the data centre and the cloud. The Altair PBS Works suite includes workload manager Altair PBS Professional, designed to improve productivity, optimise utilisation and efficiency, and simplify administration for clusters, clouds and supercomputers, from the biggest HPC workloads to millions of small, high-throughput jobs.

Appentra's Parallelware Trainer and Parallelware Analyser are novel tools for the development of C/C++/Fortran parallel code for multicore CPUs and GPUs using OpenMP and OpenACC. Designed in collaboration with experts in parallel programming from HPC, they provide a systematic, more predictable approach that leverages parallel programming best practices, and enables writing parallel codes performing to as high a standard as those written by experts.

Arm Allinea Studio is a complete suite of tools for developing Arm-based server and HPC applications and Arm Neoverse processors. It provides commercially supported products for simple and efficient code generation, and full capability for understanding and optimising workloads for optimum performance on Arm.

> "Debugging and code optimisation tools can help to reduce issues that may arise from software development"

Bright computing's Bright Cluster Manager software offers an integrated solution for building and managing HPC clusters that reduces complexity, accelerates time to value and provides enormous flexibility.

Bright Clusters for Hybrid Cloud automates the process of building and managing a Linux cluster.

Charmworks' Charm++ is a mature, highly scalable parallel programming framework that enables applications to effectively scale regardless of the execution landscape. Charm++ provides a C++ type of interface that works with Fortran, C and C++ codes. Charm++ helps creators of compute-intensive software deliver better shared and distributed memory parallel application performance.

TrinityX, from **Clustervision**, is designed to provide all services required in a modern HPC system, allowing full customisation of your cluster installation.

With TrinityX, built-in services like OpenLDAP, SLURM, Zabbix, NTP and many more

Altair

FEATURED PRODUCT

Accelerating Scientific Innovation with Altair High-Performance Computing (HPC) Solutions

The industry's most demanding scientific computing requirements, in small, mid-sized, exascale and cloud computing centres, rely on Altair to accelerate innovation and keep infrastructure running smoothly. Altair is the industry leader in high-performance, high-throughput computing solutions that orchestrate, visualize, optimize, and analyse your most demanding workloads.

Your compute journey starts in Altair Access with support for desktop integration and web portal job submission, monitoring and remote visualization. End-users are empowered to focus on science and solutions and need no knowledge of typical HPC operating environments to run simulations.

Best in class administration portals in Altair Control simplify system management, scheduling policy optimisation, user quota accounting and cost-efficient automatic cloud bursting of excess workloads.

Altair PBS Professional is the heart of the Altair HPC workload management solution. Powerful and customisable for system administrators and performant for end-users, Altair PBS Professional should be your first choice for HPC system administration and scheduling.

Limited Availability Free Trial
https://web.altair.com/altair-pbs-professional-free-trial

More information
https://www.altair.com/hpc-cloud-applications

Perforce

FEATURED PRODUCT

TotalView by Perforce

Advanced HPC Debugging and Analysis

TotalView is debugging software designed for complex applications and extensive parallelism and concurrency. It provides the specialized tools you need to quickly analyze, debug, and help find problems in your high-performance computing (HPC) applications.

With TotalView, teams can spend more time focused on the HPC solutions needed to answer life's bigger questions – and less time on debugging.

Reverse debugging enables developers to quickly solve difficult issues. They can easily record and deterministically replay the execution of programs under the debugger's control. And with the output from TotalView's memory debugging feature, you can easily find memory leaks and ensure that there are no memory errors.

TotalView's comprehensive toolkit includes debugging multi-thread and multi-process applications written in C, C++, Fortran, and Python applications calling into C/C++. TotalView supports debugging a large variety of HPC technologies including OpenMP, MPI, and debugging CUDA applications running on NVIDIA GPUs.

Visit totalview.io

are provided through standard installations. TrinityX also includes optional modules for specific needs, such as an OpenStack cloud, Docker on the compute nodes and the ability to partition a cluster.

GigaIO-powered composable infrastructure provides the freedom to disaggregate and recompose your rack resources. From CLI (command line interface) and RedFish interfaces, to complete Slurm integration and automated resource scheduling features with Quali CloudShell, or to running containers and VMs, the user can choose the tool best suited for their environment.

GitLab is a complete DevOps platform. With GitLab, you get a complete CI/CD toolchain in a single application. GitLab is a single application that spans the entire software development lifecycle. Moving to a single application can speed up your workflow and help you deliver better software, faster.

Q-Leap Networks' Qlustar is a Linux HPC/storage cluster OS that is open source, free, simple to setup, extend, operate, monitor and update. It can turn networked servers into a powerful compute cluster running simulation codes and/or a fast and scalable parallel filesystem.

ScaleMP's vSMP ClusterONE enables users to consolidate servers into a single system. An HPC cluster of up to 32 servers can be turned into a single, easy-to-manage system, eliminating all cluster complexity and reducing management overhead. The software effectively creates a super-sized workstation to enable the leap from desktop computing to HPC-scale deployment, with no special knowledge needed.

Modern companies are increasingly challenged to adapt to rapid data growth and real-time processing requirements. Xelera Suite, from **Xelera Technologies**, is a hardware-agnostic, on-demand scalable and easy-to-deploy acceleration platform which takes care of the customer's acceleration needs in a click.

Zapata's Orquestra is a workflow-based, unified toolset for applied quantum computing that allows users to compose and run quantum workflows across a range of devices, both quantum and classical, on a unified quantum operating environment.

Orquestra enables enterprise teams to orchestrate across quantum computing frameworks to compose, run, and analyse quantum workflows at scale.

Tech focus: cloud

THE LATEST CLOUD COMPUTING TOOLS AND SERVICES

Cloud computing provides a platform for scalable and quick access to high-performance computing infrastructure that can be beneficial for organisations that either need to scale quickly, or that do not want to invest in computing hardware.

Quick access to HPC-like infrastructure can deliver speed and agility for critical workloads, or enable organisations to accelerate the time to deliver scientific insights.

Cloud computing can also be used to manage infrastructure investment by enabling organisations to under-provision internal resources which can then be supplemented by using cloud resources in times of high demand. This can reduce the total cost of ownership (TCO) of computing resources by using internal resources more effectively.

As cloud services have matured, they now provide access to the networking, storage and compute resources which are a prerequisite for HPC workloads.

With the use of bare metal resources or carefully constructed virtualisation technologies, cloud providers have created a large variety of services which cater to different performance and cost requirements.

Commercial products

HPCFlow is **Advania Data Centers'** scalable AI and HPC cluster infrastructure solution. It is dedicated to customers looking to execute their AI and HPC cluster operations faster and more efficiently.

Based on Alibaba Cloud infrastructure, Alibaba Cloud Elastic-HPC (E-HPC) is an end-to-end public cloud service. E-HPC is a cloud platform providing an all-in-one HPC computing service. E-HPC provides individual users, education and research institutions, and public institutions with a fast, elastic and secure cloud compute platform that interconnects with Alibaba Cloud products.

AWS provides high quality elastic and scalable cloud infrastructure to run HPC applications. With virtually unlimited capacity, engineers, researchers and HPC system owners can innovate beyond the limitations of on-premises HPC infrastructure.

Additionally, with access to a broad portfolio of cloud-based services such as data analytics, AI and Machine Learning (ML), you can redefine traditional HPC workflows to innovate faster.

Atos Digital Hybrid Cloud is the foundation to enable, manage, and empower a multi-cloud vision and digital business. Atos Digital Hybrid Cloud combines flexibility with fast time to market, a scalable, consistent future-proofed solution and enhanced visibility and control over complex cloud environments.

The **CPU 24/7** Engineering Cloud provides flexible, pre-configured and 'ready-to-use' HPC resources, including the right CAE application to run your simulations much more quickly and securely. CAEaaS HPC systems provide high-performance servers, specialised high-throughput network components, as well as simulation application software, complex configuration management and comprehensive security measures to protect important intellectual property.

DUG McCloud is a customer-focussed, collaborative cloud solution that allows users to mix and match high-performance compute as a service (HPCaaS), professional services, and software.

DUG HPCaaS provides integrated HPC, storage, software and services, high bandwidth and low latency networking, as well as huge data storage capacity and archiving across secure, dedicated nodes with no virtualisation.

HPC is all about scale and speed. **Google Cloud**'s flexible and scalable offerings help accelerate time to completion, making it easier to convert ideas to discoveries, and inspirations into products.

Google Cloud enables

Anna Fedko/Shutterstock.com

> "As cloud services have matured, they provide access to the networking, storage and compute resources which are a prerequisite for HPC workloads"

HPC users to access custom machine configurations to control costs with committed and sustained use discounts. Build a supercomputer using the latest Intel and AMD processors, Nvidia GPUs, and high-throughput, low-latency object and file storage.

Gompute delivers solutions for HPC as a Cloud service with bare-metal compute nodes for your CAE workloads, as well as on-premise and hybrid cloud solutions.

Founded in 2002, Gompute helps CAE users to run compute intensive simulations with high demands on performance, throughput and ease of use, combined with support services.

Hydro66 offers green cloud infrastructure, providing a low emission, cost-effective and open alternative for enterprise cloud and colocation. H66cloud claims to be the greenest cloud in the world and the firm expects it to set a benchmark for sustainability. It layers an exceptionally flexible, cost-effective cloud service on top of its award-winning efficient data centre.

Inspur InCloud Platform is based on the concepts of openness, convergence, security and intelligent implementation. The platform integrates cloud computing, big data infrastructure, and capabilities through an integrated and flexible IaaS and PaaS delivery model. It helps customers rapidly build smart, highly efficient, user-friendly and stable digital infrastructures for increasingly complex business environments.

Microsoft Azure CycleCloud simplifies HPC cluster configuration and easily deploys validated, enterprise-ready reference architectures for common workload management across many industries. Deploy entire clusters, including networking, storage, compute and the entire software stack, with just a few clicks.

NetApp and **Spot** have joined forces to help companies make applications and cloud infrastructures work together seamlessly. The result is improved performance, increased agility, and up to 90 per cent reduction in cloud infrastructure costs, whether using Microsoft Azure, Google Cloud or AWS.

Jarvice is the platform that powers the **Nimbix Cloud**. Built for today's most demanding workflows, code runs faster on Jarvice thanks to Nimbix's innovative, patented Reconfigurable Cloud Computing, which runs containers on bare metal, instead of virtual machines.

Oracle Cloud Infrastructure (OCI) is an IaaS that delivers on-premises HPC power to run cloud-native and enterprise companies' IT workloads. OCI provides real-time elasticity for enterprise applications by combining Oracle's autonomous services, integrated security and serverless compute. Available for public cloud.

Red Hat Cloud Suite is a combination of tightly integrated Red Hat technologies to build a cloud infrastructure, develop cloud-native apps and orchestrate deployments across hybrid IT environments. Providing all the tools to create and be productive in a hybrid cloud.

UberCloud is a worldwide operating cloud services provider of Cloud With One Click – a secure, on-demand, browser-based self-service, fully automated and customised platform helping engineers achieve 10 times productivity for their complex engineering simulation workflows. Partnering with leading software companies and cloud providers, Ubercloud accompany manufacturers through their whole process of cloud discovery, proof of concept, and the final private/public/hybrid cloud production environment.

Xtreme-D makes HPC computing access easy, fast, efficient and economical. Their flagship product, Axxe-L, provides turnkey access to high performance systems, automating the access and configuration of HPC-optimised resources on the fly in a cloud environment. Its subscription-based usage model provides unique flexibility and freedom of access to all users.

Enabling global research through the cloud

Jisc's Matthew Dovey discusses the importance of cloud computing and improving digital infrastructure

Increasingly, research must handle more complex data at higher rates of production and consumption. Researchers require increasing computing power and sophisticated tools to help share, process and analyse data, across all disciplines – not just those traditionally associated with high-performance supercomputing.

Improved digital infrastructures are key to supporting researchers in their work. Cloud-based resources are an important element of these infrastructures and provide benefits such as increased collaboration. For instance, if multiple researchers from different institutions or even different countries are working on the same project, access to shared resources is essential. When resources are stored on a cloud platform, researchers can access these resources through their institution's credentials, from wherever they are in the world.

European Open Science Cloud

The European Open Science Cloud (EOSC) has the vision of providing a cloud platform to offer 1.7 million European researchers and 70 million professionals across science, technology, the humanities and social sciences an open service for storage, data management, analysis and reuse of research data.

The EOSC was first proposed in 2016 as part of the European Cloud Initiative to build a 'competitive data and knowledge economy in Europe'.

> "Researchers can explore and experiment with new tools, technologies and computing platforms quickly, without large upfront financial and infrastructure costs"

The aim of the EOSC is to work across borders and disciplines by federating existing data infrastructures dispersed across the EU, and will be a key pillar of the new European Data Strategy. Jisc, alongside many other partners across Europe, has been participating in the projects and working groups defining the architecture and governance of EOSC.

This approach to open research is proving particularly useful during the current climate: led by the European Bioinformatics Institute (EMBL-EBI), the European Commission has launched European Covid-19 Data Platform explicitly for research into the disease, which will be a key component of the EOSC.

The platform contains data on coronaviruses, including samples of Covid-19, which is regularly updated from international databanks. The cloud

whiteMocca/Shutterstock.com

platform also provides some tools to aid identification and analysis of the samples, which means that researchers can run experiments virtually, without having to install these tools on their own machines.

Proceed with caution

As various providers are offering more tools as Software-as-a-Service (SaaS), the role of cloud is expanding beyond that of just computing and storage solutions. Researchers can explore and experiment with new tools, technologies and computing platforms quickly without large upfront financial and infrastructure costs. That said, as with all digital tools, the benefits are dependent on how they are used.

A certain amount of caution is needed in the case of pre-built toolkits, as they can also be misused or introduce unexpected biases. If a researcher knows how to use the toolkits and understands the parameters and limitations of the technology in question, such as AI, they can be really useful. However, if this isn't the case and the tools aren't used properly, results could be compromised, for example, mistaking correlation for causation, or misinterpreting artefacts inherent in the tools as trends in the underlying data. The "black-box" nature of some of these SaaS tools – whereby their internal workings remain hidden – introduces challenges for effective peer review. While one of the advantages promoted for SaaS is that software updates are managed for users transparently, it can impact the reproducibility of research results.

> "The cloud platform also provides some tools to aid identification and analysis of the samples, which means that researchers can run experiments virtually"

However, when used well, SaaS tools can improve the flexibility and efficiency of research projects. For example, if a researcher has an idea for an experiment but is unsure whether it will work, they can spin up a machine in the cloud, instead of having to order high performance computing (HPC) kit, which can be both expensive and time-consuming. When the experiment has been modelled in the cloud, however, it can be more cost-effective to use dedicated HPC for ongoing projects. Working with accredited partners such as Jisc can also help make the most effective use of cloud services including AWS and Microsoft, further extending budgets.

It's all about balance

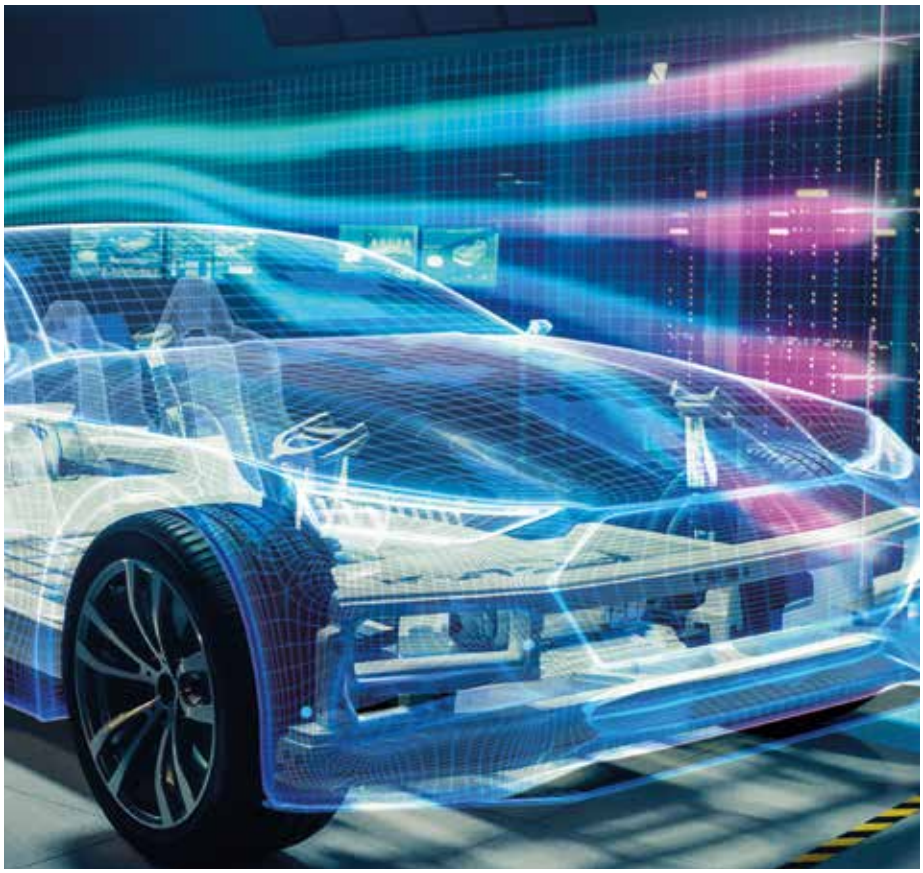
Increased availability of cloud computing for research provides options for different projects, and in the current health crisis, it is proving an important tool in global vaccine efforts. This is evident here in the UK through projects such as the University of Bristol's use of cloud to develop the synthetic ADDomer vaccine.

The global challenge brought about by Covid-19 has shed greater light on cloud's place in research collaboration, and as data analysis continues to take centre stage in research, we will only see cloud usage grow.

Matthew Dovey is head of e-infrastructure (strategy) at Jisc, and a member of the EOSC governance board. For more information on cloud solutions for education and research, visit **jisc.ac.uk/cloud**

Engineering drives new simulation tools

The increasing complexity of simulation workloads is driving a need for HPC resources in computer-aided engineering, finds Robert Roe



Gorodenkoff/Shutterstock.com

As increasingly competitive markets tighten product development cycles, the need for innovation rises. Many organisations also want to reduce expenditure and pool resources to deliver a more efficient computing infrastructure, which is increasing the use of cloud and threatening the traditional general-purpose HPC cluster.

These factors are driving simulation engineers across several industries – from traditional HPC users such as aerospace and automotive, to more recent adopters, including construction, robotics, autonomous systems and healthcare – to turn to HPC simulation. Organisations need to increase capacity for simulation, while delivering a better-optimised platform that can be part of a strategy to reduce costs.

The need for HPC

The rising use of digital twins to shorten product development and reduce the costs of validation and testing is also increasing the demand for computational simulation resources. Today, engineers need high-performance computing and large scale simulation so they can deploy new materials, refine their understanding of composites or to better understand physical phenomena such as combustion or turbulence, with more fidelity than was previously possible.

This challenge is compounded by the need to push the boundaries of simulation, across CFD, materials physics or interactions between components and systems, which also drives the need for more complex and larger simulations. The challenges of developing innovative products in shorter timespans increases the pressure on organisations to deliver more from each development cycle.

All these use cases are driving a need for more complex and higher fidelity simulation across the majority of verticals engaged in computer-aided engineering (CAE) simulation. While traditionally HPC may have been employed by oil and gas, automotive and aerospace users, today it proliferates across all verticals in CAE).

A Grand View Research report in February found that 'the global CAE market size was valued at $7.3bn in 2019 and is expected to register a CAGR of 9.3 per cent over the forecast period' (2020-2027).[1]

It also notes that the industry is shifting from on-premise to cloud computing in order to reduce costs of hardware provision and software licencing.

However, these same market drivers are also influencing the development of pre-configured simulation platforms, which are effectively HPC appliances that are specifically designed and optimised to run HPC CAE simulations with common ISV applications.

Intel Select Solution[2] portfolio are pre-configured and benchmarked computing solutions designed for a variety of computing applications, including HPC. The modelling and simulation variant, delivered through a partnership programme with several HPC integrators, is designed for CAE simulation users that want to accelerate infrastructure deployment.

The solutions are designed to overcome challenges with hardware selection and software performance to ensure optimised engineering simulations. This enables organisations to scale quickly with a workload optimised solution that contains the latest HPC, a high bandwidth fabric and fast storage.

Increasing efficiency

HPC provides performance but this comes at a cost of infrastructure and expertise to get the most out of a cluster.

Without the correct optimisation and staff knowledge, an organisation will

not be able to maintain the system or to deliver the required performance from their applications.

For businesses that do not have the internal infrastructure and staff to manage a cluster, or startups and small firms that want to bypass that investment, or want to adopt a hybrid approach to managing compute resources, pre-configured HPC solutions tailored to simulation can combine performance and usability.

Compute-intensive simulations found in today's industrial design projects require computational resources that can deliver modelling and simulation workloads at scale. To support those applications, modern HPC systems require multi-core processors, high-bandwidth fabrics, fast storage and sufficient I/O capabilities to not bottleneck the performance of a given application.

To deliver this kind of hardware, typically an organisation would need staff or additional help to build a balanced architecture to support the specific set of applications.

While this is a complex process, there is a risk of buying sub-optimal hardware, then wasting a potentially expensive component that cannot be fully utilised because of a memory or I/O bottleneck

"While buying a general purpose solution may deliver the performance an organisation requires, it may not be the best solution for users with specific CAE application requirements"

for example. The complexity and variety of technologies available on the market mean that designing and assembling an HPC system for specific workloads can be time-consuming. This is particularly true if there are a variety of applications that need to be benchmarked or validated with a new architecture.

Pre-configured solutions, such as the Intel Select Solution, reduce the burden of designing, provisioning a cluster that must then be optimised and maintained to deliver consistent service for the users. Employing these technologies provides the peace of mind that your application will run effectively and deliver the performance required without large-scale optimisation, because it is designed to run ISV codes commonly used in simulation today.

Whether an organisation wants to deploy cloud or in-house resources, build its own cluster or buy a pre-configured solution, the total cost of ownership must be considered.

While buying a general purpose solution may deliver the performance an organisation requires, it may not be the best solution for users with specific CAE application requirements. Efficiency, infrastructure costs, staffing and electricity bills can all drive decision making to a reduction in the total expenditure on computing infrastructure.

To deliver the most efficient solution possible, organisations can look to deploy a cluster which has been specifically built and optimised for CAE applications. These systems can be tested ahead of time, which means that engineers can understand how fast their real-world workloads will run on a specific hardware configuration.

References

1. www.grandviewresearch.com/industry-analysis/computer-aided-engineering-cae-market
2. www.intel.com/content/www/us/en/products/solutions/select-solutions/hpc/simulation-modeling.html

Real-life problems with big data solutions

Big data is helping research into patients with sleep problems, reducing trips to medical facilities, and assessing their safety behind the wheel

Ian Smith heads up the sleep service at Cambridge's Royal Papworth Hospital in the UK, which is contributing healthcare data for the Track and Know project, aimed at finding big data solutions to real-life geospatial problems.

Royal Papworth accepts referrals from across the East of England; an area of 19,000km^2, with around 20,000 patients seen each year. The population is thinly spread, with an average density of 310 people per square kilometre – compared to more than 1,500 in London.

Smith said: 'We often need to arrange sleep studies for patients at home that would require several trips to and from the hospital, which would add up to journeys of as much as 400km for some people. The solution has been to develop a network of outreach facilities over the last 10 years to improve accessibility, but these have been situated as best guesses to serve our referred population. At the peak we had 23 facilities and the average journey was 35km.

Through the project the team – a mix of 14 university and commercial partners from 10 EU countries – has analysed 46,211 historical patient journeys and modelled the most efficient network.

Smith continued: 'The model shows that we can provide the service most efficiently with just 10 clinics and reduce the average journey to 17km, saving up to 30 tonnes of CO_2 emissions a year.

'We have looked further into our referral base by the socio-economic status of the areas from which patients are referred; we would anticipate more sleep illness in areas with social deprivation related to poor employment opportunities, increasing rates of obesity, and increased numbers of older residents.

'However, we have had fewer referrals from these areas. When patients do attend they are more unwell and a higher proportion of patients fail to attend their appointments. We also now know that more people miss their appointments when it is sunny than when it is raining! We are re-running these high-level, complex models to see if we can make journeys particularly easy for people referred from areas that are socio-economically challenged.'

vchal/Shutterstock.com

> "The plan is to use the app to alert people when their driving performance is drifting"

Driving knowledge

The team is also working on a sub-project that examines driving behaviour among those with sleep disorders.

Smith said: 'We know that people with sleep disorders are more likely to crash their cars than people with good sleep, maybe four times more likely. It also seems likely that shift workers are also at risk. However, there are currently no good measures of who is at risk and who is safe.'

Accordingly, the team is developing a mobile phone app to measure driving behaviour, and is currently monitoring the driving of staff at the hospital who work a mix of day and night shifts. It will soon start monitoring patients with sleep illnesses, before and after treatment, as they drive around the East of England region.

Smith said: 'The plan is to use the app to alert people when their driving performance is drifting. In future, it should be able to help adjudicate on whether or not people can return to driving when they have had a sleep illness – the aim is to contribute to an improvement in road safety, and also getting people back to work as soon as it is appropriate.

'Other use cases for the Track and Know platform include looking at insurance models based on driving behaviour, and fleet management – predicting journeys and efficient distribution of resources.

'The data analysis teams have built the Track and Know platform to accept a broad range of data types and allow for rapid processing of big data, integrating driver behaviour with local road conditions, weather patterns, distribution of retail outlets, petrol stations, electrical charging points, weight of traffic, known accident black spots, and so on.'

The study is funded from an EU 2020 grant of €4.8m, and all publications relating to the research are required to be open – they can be found at https://trackandknowproject.eu/publications/. Furthermore, many of the models and algorithms are based on the EU Track and Know project, where some of the data and models are public: https://trackandknowproject.eu/file-repository/

Smith concluded: 'National Health Service journeys account for around 5 per cent of traffic on our roads. If we can reduce this, we can have a major impact on road congestion and CO_2 emissions.

'In the era of Covid there is an increasing premium attached to reducing travel and visits to healthcare facilities unless absolutely necessary. Sleepiness is now a bigger problem for road safety than alcohol, so identifying sleepy drivers is a key priority and a simple phone app could give drivers their own safety profile.

'Replanning services throughout the NHS may help clear backlogs caused by the lockdown, and better match supply and demand in the coming years. On the back of the current work, a sub-group from Track and Know has been awarded a grant from Innovate UK to address exactly this challenge. Watch this space!'

FREE ON-DEMAND WEBCAST*
AVAILABLE NOW

Scientific computing in the era of AI

SPEAKERS

Carlo Ruiz
EMEA Director - AI Data Center Solutions, Services & Operations
NVIDIA

Timothy Lanfear
European solution architecture and engineering team manager, Enterprise Solutions Group
NVIDIA

Dive deeper into how NVIDIA can speed up your research - across healthcare, weather forecasting, seismic mapping, and more. From data analytics, to simulation, AI, and visualisation the NVIDIA platform is built to accelerate scientific workflows, end-to-end. NVIDIA powers 8 of the top 10 supercomputers in the world and is focused on getting data scientists and developers the tools they need to achieve their life's work.

Join us for this webinar and get acquainted with NVIDIA's scientific computing offerings, including a closer look at NVIDIA DGX A100 and how NVIDIA is helping researchers and scientists in their fight against COVID-19.

In this webinar you'll explore:

- How NVIDIA can accelerate your research and enable breakthrough research
- The next wave of supercomputing and its impact
- Strategies for successful AI adoption in a variety of sectors
- The ways in which NVIDIA is supporting scientists and researchers as they race to find a cure for COVID-19
- Specific tools and resources to support your AI projects

SCIENTIFIC COMPUTING WORLD

NVIDIA.

www.scientific-computing.com/webcasts

*This webcast is FREE to view, though registration is required

Transforming the laboratory

SOPHIA KTORI DISCUSSES THE NEED FOR DIGITAL TRANSFORMATION IN THE LABORATORY

Specialist consultancy firm Astrix Technology Group works with life science organisations to identify and implement systems for scientific informatics and laboratory operations, to maximise value from scientific data, improve product and experimental quality, and aid safety, security and regulatory compliance.

Dale Curtis, Astrix president, said: 'We're not a technology company and we don't make software, but that's a big advantage for us because we can take a vendor-neutral stance and really help companies to look at their operations and workflows, and identify just how they can start to approach digital transformation.'

Sometimes that transformation will start from a very basic position of an initial move away from pen and paper, and sometimes from the point of acknowledging that their legacy informatics systems are no longer viable.

Review and be realistic

Whatever the ultimate aim, any attempt at digital transformation should start with a review of existing processes, and consider what the organisation hopes to realistically achieve, in terms of optimising those processes.

'The goal may be to remove the need for manual data entry, but to do that you have to understand your processes and your process flow, to understand the end-to-end workflow,' Curtis said. 'Organisations need to take a tactical approach based on need, rather than just focus on the software that's out in the marketplace.'

Astrix has built its expertise on many years of researching laboratory best practices and implementing solutions for scientific and laboratory-based organisations. 'One starting point can be to measure just how much time scientists spend on activities such as the manual transcription of data from one system to another, data review verification, reports preparation, extraction of data,' Curtis noted. 'Digitising some of all of these activities can dramatically improve efficiency and productivity, as well as help to ensure data integrity and improve downstream quality. But you have to go through your processes to understand how that digitisation needs to be implemented.'

In reality, companies are often quick to flood laboratories with software, he suggested. 'They may have a dozen or more informatics systems, including multiple LIMS and ELNS from different vendors, which don't talk to each other. That causes another set of problems.'

Consider bite-size chunks

Companies commonly approach Astrix because they want to understand how digitalisation can deliver value back to their organisations in a commercially reasonable timeframe, Curtis suggested.

'They may have been told by vendors that digital transformation will eventually transform the way that business is done, but that it will be a three-year initiative and cost however many million dollars. That's just not tenable for most organisations.'

It's likely that the most successful projects won't be those that promise an all-encompassing, enterprise-wide outcome, he added. 'It is generally more manageable to structure digital projects in bite-sized chunks that can be achieved in a realistic timeframe, rather than trying to boil the ocean.'

Take this approach and its possible to address end-to-end processes within a lab operation one at a time, however small a portion of overall workflow those processes are. 'We could start looking at the beginning, or we could start in the middle,' depending on where the major sticking points are,' Curtis noted.

Astrix's approach is to help companies define their processes, then use its expertise and technology selection process – which over the years has encompassed potentially hundreds of thousands of requirements – to understand those processes and define potential solutions. 'Then we can also start working with the vendors themselves, to demonstrate how their product or platform can deliver against specified objectives.'

Beaten by complexity?

Astrix is vendor-independent but, as Curtis pointed out, the company is frequently called in because a vendor hasn't been able to deliver on a promise. 'We carry out upwards of 200 projects a year, and more than half are rescue projects, perhaps because the vendor has started an implementation and realised that it's a lot more complex than they thought, or the solution isn't really addressing the finer details.' And as technology becomes more 'capable', then the differentiation between acronym-led decisions, 'Do we need a LIMS, ELN, SDMS?' starts to disappear, he continued. 'Companies shouldn't just think about platform 'names'.'

'Once there is an indication of how the different steps can be digitised, and what that digitalisation will achieve, we can bring in the architectural elements, and look at the systems or platforms that might achieve that architecture, and how they can be used to improve and streamline

Gorodenkoff/Shutterstock.com

defined processes, for example to improve security, compliance, reduce manual errors or save time.'

Then it is possible to look at how software can do that in a commercially reasonable timeframe. 'No one wants to sit around and look at projections and implementation plan PowerPoint slides for a year or more. They want to see progress in the short term, and demonstrate real value driver points in three to six months.'

Change and cultural challenges

Another major consideration is the change in working practices that digitalisation will inevitably engender. Whether that's moving away from the reliance on paper, pen and excel spreadsheets, or replacing existing, but outdated legacy systems with modern platforms. 'In fact, some of the biggest challenges aren't technical at all, but cultural,' Curtis noted.

'You're asking a lab scientist who has been trained to execute a method, design an experiment, or write a report in the same way for potentially decades, to change their working practice from the ground up. That's quite an expectation. Even if you have technical and supervisory support, it's important to clearly define a value-engineered roadmap that shows each individual how, by changing their behaviour, they are going to impact on their own working practices, and the overall process.' But again, do it in bite-sized chunks, and it becomes less daunting. 'Everyone needs to understand the benefits in the long run, and people are just as much part of that transformation and journey as the software.'

> "It is generally more manageable to structure digital projects in bite-sized chunks that can be achieved in a realistic timeframe, than trying to boil the ocean"

Not trying to change everything in one go – like the entire analytical operation of an organisation – also makes the process less intimidating from a user perspective.

'Maybe start with a single laboratory that is part of that analytical engine in an organisation. Take this one step further back and you may even want to identify an analytical step within that lab, perhaps.

'For example, how does a chemist submit essential material for characterisation? There you have that easy-to-manage chunk but it's still a workflow and you can map it out from end to end. How do the samples come in? How do the materials go through the workflow? What do we produce to then send on to the next group? By looking at these processes in stages, we can define what each package of data actually looks like, and that's a really digestible piece of a problem … you're not trying to eat the whole elephant.'

And this user experience leads to the issue of training, Curtis said. 'It's all very well implementing a new system, but how do you ensure that everyone knows how to use it and is comfortable doing so? One thing that the pandemic has demonstrated is that there is no longer the need to bring everyone into a classroom over multiple days and try to teach by rote in an external environment,' he suggested.

People need to be confident in their own workspace, he continued. 'Smart companies are investing in learning management systems, so the people using a new software or informatics platform can continue to learn as they are using it, and become confident in the workplace, which is, after all, where they are using it from day to day.' →

→ **What works, and what doesn't?**
Life science enterprise laboratory informatics provider Benchling also notes that companies looking to either initiate or update their digital transformation journey should start by taking stock of their existing infrastructure. Michael Schwartz, head of product marketing, said: 'Organisations must be honest about what's working and what isn't, and also about what they want to achieve, rather than focus on specific categories of software. It really has to do with their workflows and R&D goals.'

Complexity aside, companies are still concerned about some of the fundamentals, such as security. Every organisation is in a different spot in their transition into the cloud, Schwartz noted. 'Most companies will still need assurance that their data will be as secure, if not more, in the cloud than it is on-premise.' Connectivity is another key consideration, according to Schwartz, and the ability to integrate platforms from different vendors comes high on the list of requirements.

'Organisations are aware that they will, if they don't already, implement multiple platforms across their organisation – not just in their R&D.'

So while their existing processes may have little connectivity, that ability to integrate different platforms – whether those already in place or those yet to be purchased – is also high on the list of digital requirements. 'Organisations know that if their data doesn't aggregate in a very clean and structured manner, they are going to have challenges. They want to make sure that they have that connectivity, so that they can integrate across applications and data repositories.'

Outdated platforms
A willingness to look at existing software objectively will highlight problem areas with possibly outdated platforms, Schwartz noted. 'Early generation ELNs, for example, offered a fairly narrow field of data capture. They were designed primarily to capture unstructured data, and to document intellectual property.' But the software was not well connected as part of a holistic R&D operation.

In fact, such platforms were not much different than a paper notebook, he suggested. 'Individual scientists received access to an ELN for instance, but there was no connectivity, so any experimental data that was put in the ELN wasn't readily accessible for sharing and scientific decision support.

'That's not the way science works now. There may be multiple specialised teams all working around one common endpoint,


Gorodenkoff/Shutterstock.com

"It often starts with a need to document molecular and experimental design in the same location, and how their scientists can achieve individual productivity gains"

and they need to organise around data generated by multiple individuals and different teams to drive them to that endpoint.'

Flexibility and connectivity
Since Benchling was founded eight years ago, Schwartz said it has developed a flexible, unified, cloud-based informatics solution for life science R&D organisations, ranging from the largest enterprise pharma companies to the most cutting-edge biotech startups.

'Our modern approach and scientific expertise has allowed us to build a highly flexible platform which can be easily configured to support a diverse set of R&D functions in this space, which are all linked through a connected platform. We can now address some of the mission-critical pain points in life science R&D. Individual organisations all have their own distinct challenges', Schwartz said.

'These may centre on managing inventory, or on tracking samples, or on centralising and standardising data capture, but the important thing is that we can address all of these different use cases on one unified platform.'

Companies in the biotech and pharma industry are at various points on their digital transformation journey, Schwartz stated. 'In life sciences, it often starts with the need to document molecular and experimental design in the same location, and how their scientists can achieve individual productivity gains.

'Then companies typically want to address operational inefficiencies, which usually involves rethinking how they standardise their database of biological and chemical entities, and how they co-ordinate requests and workflows.

'Lastly, companies will then seek to drive smarter program-level decision making, which involves aggregating their data for analysis and visualisation. Often these different functions are addressed holistically at the same time, but it's possible to solve for these acute needs in phases, as part of a broader transformation process.'

Benchling believes two key factors set it apart from some other platform providers in the same field. The first is that all the implementation teams that work with clients have deep domain expertise in the life sciences, enterprise software, and data modelling, so they understand the scientific needs of biotech and pharma organisations, as well as the IT and business requirements.

'Whether an organisation is focused on small molecule R&D, or is involved in the biologics arena, they will work with a Benchling team that has carried out multiple implementations in their area of specialisation, from traditional chemical drug discovery, to vaccine research, gene therapy or synthetic biology and industrial applications,' said Schwartz.

The second hallmark of the Benchling solutions is that the platform can be implemented without code. This makes the platform fast to configure and easy to deploy. 'We run an agile implementation process, and iterate very quickly with our customers, so we can get a prototype up and running quickly. This can then be fine-tuned so that customers get maximum benefit, without complex customisation processes.'

The company has developed a mature set of Rest APIs, so connecting the Benchling platform to existing systems should not be problematic. 'But we also have partnerships with instrument vendors, and have developed templates for common laboratory instrumentation, so most clients can get up and running, and connect with almost any instrument very quickly,' said Schwartz.

'As the aim is to automate those manual processes as much as we can, we have a lab automation module that sends run instructions to robotic liquid handlers and analytical instrumentation, then digests the output data in a structured, actionable format.'

Subscribe for free*

SCIENTIFIC COMPUTING WORLD

Do you compute?

The only global publication for scientists and engineers using computing and software in their daily work

Do you subscribe? Register for free now!

*Registration required

scientific-computing.com/subscribe

Also published by Europa Science

Covid-19 accelerates the adoption of remote working tools

ROBERT ROE LOOKS AT HOW COVID-19 IS DRIVING CHANGES IN THE LABORATORY, INCLUDING TOOLS FOR REMOTE WORKING AND HUGE INCREASE IN DEMAND FOR TESTING COVID-19 SAMPLES

The pandemic is increasing demand for testing, genomics studies and vaccine and drug development. To facilitate this research, and scale up testing capabilities, laboratory scientists are using laboratory information management system (LIMS) and electronic laboratory notebook (ELN) software to help manage collaborative work, and promote reduced-capacity working environments in the laboratory.

Patrick Rose, digital science product manager at Thermo Fisher Scientific, notes that there has been a pivot in the tools used in laboratories due to Covid-19. 'The labs are starting work leaner, they are having to collaborate with other partners, other contract research organisations [CROs] to help battle the pandemic and find the vaccine.

'What we are seeing is an increased need to work remotely due to alternating days, slimmer shifts trying to minimise the time people are in there together. But, they still need to work on their results, gather and share data and collaborate,' added Rose. 'So the tooling and solutions, like the LIMS that we offer, become more pivotal and critical. It is being able to see what other users have done – potentially what they may have done in the last shift, what you need to do for the next shift.'

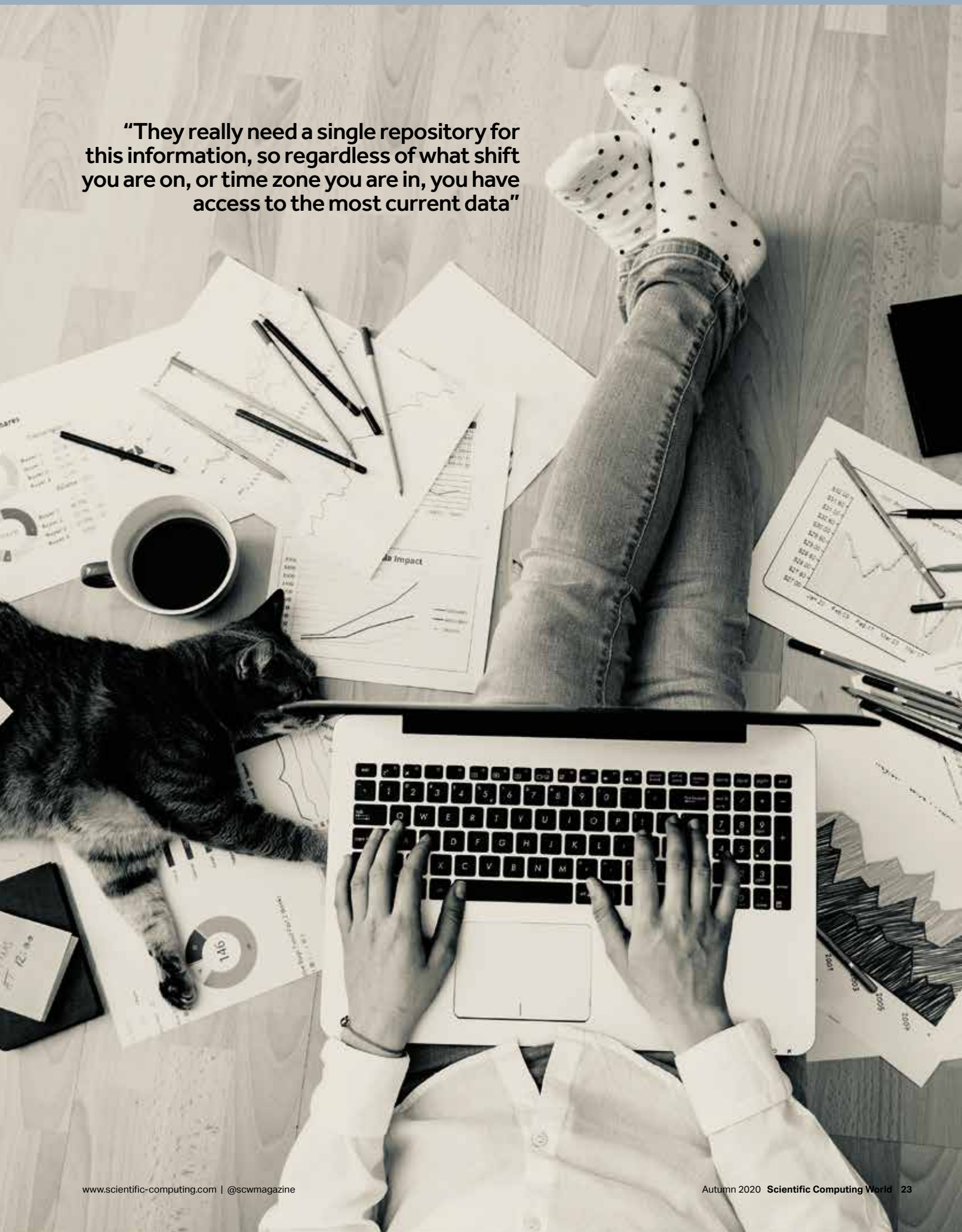
Collaboration is also playing a big part as CROs and other partners can help to take up extra workloads during peak times or enable organisations to work with disparate teams that are not using the same facilities to make up for the loss in staff numbers.

'Scientists are using tools to take the data they are using in the lab to quickly write down their results, so they can then extrapolate when they are out of the lab, or in a more separated environment. They can still access all of their information and use it,' added Rose. 'We are also seeing labs driving tighter collaborations with CROs because of everything that is going on. There is this collaboration among companies that was unseen before, where everybody is working together to try to cure this. With everything going on in the world, it's not just about the bottom line, it is about trying to help people, and get everybody back outside again.'

The increased demand for collaboration – either between organisations or just using tools to better facilitate a lean working environment for internal laboratory teams – means laboratory workflows can continue without major disruption.

Lauren Taylor, digital science solutions manager at Thermo Fisher Scientific, →

Creative Lab/Shutterstock.com



"They really need a single repository for this information, so regardless of what shift you are on, or time zone you are in, you have access to the most current data"

Rido/Shutterstock.com

noted that 'scientists in different labs are working across time zones and across countries. They really need a single repository for this information, so that, regardless of what shift you are on, or what time zone you are in, you have access to the most current data'.

Tools facilitate remote working

LIMS or ELN systems can help to minimise the amount of time scientists have to spend in the lab, as mobile alerts or other features can help notify scientists when their experiments will be finishing. Streamlining data sharing and access also helps facilitate agile decision making and lets scientists access data and analyse or work on data from home.

While some of these tools have been around in certain forms for a number of years, Covid-19 is forcing organisations to look carefully at how they manage access to the laboratory, while maintaining as high a level of productivity as possible.

'The platform for science helps to manage the end-to-end process through the labs – it enables scientists to know what they did and what someone else has done. It also gives you the ability to integrate with tools like RStudio,' said Rose. 'Labs are embracing the cloud and embracing remote working, where in the past there was some nervousness around the cloud. People are starting to open up and realise that they do not need to be in the lab. You can be outside and still get the same amount of work done.

'This shift that is happening now, because of Covid-19, is going to continue, because you are getting the same amount of work done, if not more, because you are not spending an hour a day in traffic,' Rose added. 'Automation tools are helping to drive this as well, because you do not need as many people in the lab with things becoming automated.'

Several LIMS and ELN vendors are now equipping their software with pre-configured workloads to help scientists get set up with key laboratory operations. This reduces the amount of setup and administrative work that needs to be done by scientists, which should lead to increased productivity.

> "What we are seeing is an increased need to work remotely due to alternating days, slimmer shifts, trying to minimise time that people are together"

'Folks want to sequence not just the virus itself but also the host genomics too, and how the genetics of the host may affect the healthcare outcomes. There is a lot of that sequencing information, for example, that needs to be gathered and they are pretty complex workflows,' said Taylor

'Some of those workflows we have already pre-configured to get customers up and running as quickly as possible. In terms of the integration of instrumentation, for example, you may have specimens in the lab that you want to sequence, that information needs to be put onto the sequencer so that you know which samples you are sequencing, and that information can be relayed back to the correct sample,' Taylor continued.

'The integration really is automating the push of information to the sequencers, so that you are not copying and pasting or uploading and downloading files manually.'

Testing capabilities

In addition to pre-configured workflows to assist in the genomics research into the novel coronavirus, Thermo Fisher has also released the TaqPath Covid-19 Multiplex Diagnostic Solution – a multiplex real-time RT-PCR diagnostic kit to enable clinical and public health laboratories to quickly

diagnose Covid-19 caused by the Sars-CoV-2 virus. The Applied Biosystems TaqPath Covid-19 kit is designed to enable fast, highly sensitive multiplex diagnostic testing that contains both the assays and controls needed for real-time PCR detection of RNA from the virus. This can mean detection in under two hours.

In a recent blog post from Thermo Fisher, Kamini Varma, senior director molecular biology, gave an inside view on the efforts taken to launch the TaqPath kit.

'We had already developed a respiratory panel in 2019, so we were quite familiar with coronavirus and how to design assays to identify it. That didn't make the work ahead easy, but we were able to quickly mobilise to develop both single and multiplex tests. I guess you could say that over the years we've been "building our muscle" to be able to respond quickly and decisively to a pandemic like this, and it's a good thing we were.

'By mid-February, our nearly round-the-clock development work had produced a viable assay that we felt was ready to begin the emergency-use authorisation [EUA] process with the FDA [US Food and Drug Administration], a first step in putting a test into the market for more widespread use,' added Varma. 'At the time, we were unsure of the timeframe for the EUA process, but our reference point was the Zika virus a few years ago, where our EUA process took nine months. But, given what was already happening in countries such as Italy, we knew we would need to find a way to accelerate the process.'

As scientists across the world continue to fight against this global health crisis, technologies for testing are being developed, manufactured and scaled up to meet unprecedented demand. Recently informatics software provider Abbott launched an antibody test for Covid-19 which is being shipped to hospitals across the US.

The test is a serology test – also called an antibody test – providing a critical step in battling the virus by enabling health professionals to determine who has been previously infected.

Abbott's test helps to detect the IgG antibody to SARS-CoV-2. An antibody is a protein that the body produces in the late stages of infection, and may remain for months after a person has recovered. Detecting these IgG antibodies will help determine if a person was previously infected with the virus that causes Covid-19.

Abbott first made the test available as part of the FDA's notification without an EUA pathway that was outlined for Covid-19 diagnostic tests during the public health emergency. Since then, Abbott has received EUA from the FDA and CE Mark in Europe.

In April the company announced that it would 'significantly scale up' manufacturing for antibody testing. In August Abbott reported that it had shipped more than 13 million antibody tests to customers across the US.

While molecular testing can be used to identify people with the virus, antibody tests can tell whether someone has been previously infected. This type of knowledge will enable scientists to better understand how long these antibodies stay in the body, and if they provide immunity. This information can also help public health officials understand how widespread the outbreak is and could help support the development of treatments and vaccines for Covid-19.

'Antibody testing has the potential to unlock a lot of unknowns about this novel virus," said John Hackett, divisional vice president of Applied Research and Technology, Diagnostics, Abbott. 'Having tests that can work in different healthcare settings is critical to our understanding of the virus, and to helping give healthcare providers answers they need about their patients.'

> "Allowing the scientists to do less administrative work and spend more time working with their data is going to speed up innovation"

Interactive Software's Achiever medical LIMS can be used to facilitate onsite testing for universities. In a recently published blog post Sharon Williams, product director at Interactive Software, discusses the role of LIMS software in assisting organisations to conduct their own tests for Covid-19.

'Achiever Medical LIMS formats and standardises your data to make sure it's consistent. If you're not carrying out the tests immediately and want to store the samples, storage management in the solution enables you to define your storage containers and track sample movement.

'If you're handling human tissue samples in the UK, then compliance with the Human Tissue Act is essential. So, it's important that you have the appropriate level of consent for any work you're doing with the samples. Especially if you're planning on aliquoting and storing them for future studies. Achiever Medical LIMS enables you to record and track informed consent alongside the student or staff member. This includes managing opt-ins and handling consent withdrawals. Plus, all sample events, such as aliquoting and analysing, are audited and the sample genealogy captured for complete traceability,' states Williams.

'Achiever Medical LIMS is available as an on-premise or hosted solution. The solution also includes a set of standard sample processing workflows as well as a library of dashboards. This allows your labs and Biobank to start using the system within weeks,' adds Willams.

Enabling remote working

One of the key concepts behind web-based LIMS and ELN software is the access that it provides to users. With current working conditions organisations are taking additional precautions to protect staff and tools that can facilitate remote working are becoming increasingly critical to support organisations ability to deliver results in a timely manner.

'One of the ways to expedite gaining insights from the data is the ability to quickly share data or have access to data,' notes Taylor. 'The ability to have that information at your fingertips, rather than having to wait for an email update from a colleague, or that weekly meeting update to see what the data is looking like. Now you can choose to share that data with your collaborators and generate insights closer to real time, rather than waiting for prescribed communication moments.'

Rose added: 'Being able to allow collaboration but within the confines that make sense between two organisations. Having tools that enable you to secure things just based on the information that you really want to share. When you get more remote, the need for web-based access becomes even more critical.'

Taylor stressed that ultimately these tools are about increasing productivity and minimising time wasted by scientists in the laboratory. 'Allowing the scientists to do less administrative work and spend more time looking at and working with their data is going to speed up innovation.'

Having to constantly check on instrumentation, calibration settings or reagents means that scientists' time can be wasted. Making use of informatics tools that automate some of these processes or allow the information to be easily checked in one central location can alleviate this burden. 'You can automate some of that work so scientists can spend more time with the data and come up with breakthroughs,' said Taylor.

Design automation

The role of Antha in developing a more effective platform for design of experiments

Automation offers higher throughput and reproducibility than manual R&D. However, programming automation requires large amounts of time and specialist expertise, and can result in inflexible workflows that require substantial re-coding when needs change, as they often do.

To combat this challenge Synthace have developed a platform to help automate the design of experiments (DoE) so that organisations can increase the complexity of their investigations, leading to more robust and productive experiments.

Markus Gershater, Synthace co-founder and CSO, said: 'Antha takes in much higher-level specifications of the automated protocol that a scientist wants to run, and automatically generates the script needed to carry out that protocol on the make and model of automation that the scientist has specified. This capability is based on the Antha language, which allows the logic of lab protocols to be expressed in code.

'These Antha workflows are equipment agnostic, and equipment-specific instructions are generated by Antha automatically. What this means, in practice, is that automation can be programmed extremely rapidly, and automated protocols can then in turn be rapidly adapted on demand.'

While, in principle, all manual R&D could ultimately benefit from being automated, for most organisations automation has traditionally focused on high volume workflows that require lots of repetition with little overall change.

'A good example of this is high-throughput screening, where the same assay is to be run thousands or even millions of times,' said Gershater. 'Antha's dynamic generation of automation scripts enables automation to be rapidly adapted for more changeable contexts, which encompasses much of biologics R&D. Experiments can be automated where it would previously have been impractical to do so.

'Any experimentation by hand is limited by the amount of complexity that can be handled by the scientist doing the work. This means that the experimental designs people would ideally like to run are unfeasible in practice. Antha can effortlessly generate the instructions needed for even highly complex experiments, freeing scientists to do the experiments they need to address biological complexity,' suggested Gershater.

SergeyKlopotov/Shutterstock.com

> "Experiments can be automated where it would previously have been impractical to do so"

'Automation offers huge benefits over manual experimentation: reproducibility, throughput and decreased hands-on time, freeing scientists to spend more time designing experiments and analysing results.

'Antha adds to this, introducing transferability and sharing of automated protocols across sites, even if different locations are using different equipment makes, and the ability to do experiments that were previously intractable with either manual experimentation or automation programmed in a conventional way. We have seen our users gain large amounts of additional scientific insight, while saving large amounts of time that would otherwise be spent in laborious manual work or programming.'

Optimising therapy development

One example is Oxford Biomedica, who started using Antha in 2018. Oxford Biomedica employed the Antha platform in their research and development to improve the efficiency and robustness of their in-house lentiviral vector production through optimisation of their transfection reagent mixes.

Lentiviral vectors are able to modify a broad range of dividing and nondividing cell types. This leads to stable and long-term expression of the gene or genes of interest. They are used in in-vivo gene therapy applications for central nervous system indications, such as Parkinson's disease and Amyotrophic lateral sclerosis.

To investigate this complex biological system a multifactorial experimental approach was taken that would have been very challenging via a more traditional approach using manual execution. This is largely due to the number of experimental runs and the complexity of the experimental design.

Oxford Biomedica shortlisted 10 factors for investigation in the first iteration of the DoE optimisation. This allowed for a more complex analysis of how different factors interact. The Oxford Biomedica team also investigated the benefits and value of deviating from a traditional iterative DoE methodology for a quicker optimisation strategy.

Full details of the DoE process can be found in the case study but Oxford Biomedica reported an 83 per cent time saving from combined design, planning and physical execution, and a 32 per cent resource saving.

In a webcast with Synthace: 'DoE Optimisation of Advanced Therapy Development with Antha', André Raposo automation group leader, Oxford Biomedica, noted that using the platform for DOE had 'fundamentally changed the way we work'.

'In the old days, DOE could be a bit insecure somehow. The way that is now we do not have to worry, the system does everything and the capacity is much larger, so you can do DoE's with a capacity of 200 to 500 wells. That would be impossible before,' added Raposo.

Webinar:
Oxford Biomedica Discuss DoE Optimisation of Advanced Therapy Development with Antha
https://vimeo.com/393991672/5ea57a90a3

Case study:
Design of Experiments (DoE) with Oxford Biomedica
www.synthace.com/customers/case-studies/detail/design-of-experiments-doe-with-oxford-biomedica/

Harness big data in the laboratory

Jean-Charles Wirth, Merck's head of applied solutions for life science, discusses the firm's acquisition of BSSN and the software they develop for laboratory scientists

whiteMocca/Shutterstock.com

We spoke to Jean-Charles Wirth a year on from Merck's deal to buy BSSN, to see how they are developing software products for laboratory scientists and continuing BSSN's work on the AnIML standard.

What drove the decision to acquire BSSN?
Modern scientific laboratories generate immense amounts of data daily, but the value of that data remains largely untapped if it is hidden away in silos or produced by multiple, incompatible sources.

Merck acquired BSSN Software to accelerate our customers' digital transformation in the lab. Our customers can now harmonise data across biological and analytical instruments and harness the power of big data in the lab, which is a critical component for organisations seeking to make breakthrough advances in AI and machine learning.

Regardless of the software or data systems that a company may already be using – such as ELN, LIMS, SDMS, ERP, and so on – our software can collect and convert scientific data into a single, unified format and seamlessly integrate into existing workflows.

What types of software or research tools do you provide to scientists?
Our mission is to solve the toughest challenges in life science and, as the convergence of technology and science in the lab continues to progress, we see numerous opportunities to enhance our customers' lives by solving software-related usability challenges.

Dating back to 2015, one of the strategic drivers of the Sigma-Aldrich acquisition was the strength of its digital capabilities and e-commerce platform. Since then, we have continued to make many investments in developing digital solutions for our customers. We provide a broad range of digital solutions to our customers, depending on their needs.

This ranges from laboratory management software such as the Lanexo System, to our Synthia retrosynthesis software that helps chemists analyse custom synthesis pathways, and large-scale joint ventures like Syntropy, which is designed to solve enterprise-level oncology data-related challenges. Digital solutions will continue to be a focus area for Merck in the years to come.

> "Merck acquired BSSN Software to accelerate our customers' digital transformation in the lab"

What is the Lanexo System?
Our Lanexo laboratory inventory, safety and compliance management system is a digital laboratory informatics solution that helps improve efficiency, drastically reducing the time that researchers need to spend on repetitive, error-prone tasks. We estimate that the system can provide a 97 per cent time saving on inventory management and a 92 per cent saving on setting up an experiment and documentation, as well as improving data quality and traceability.

The Lanexo System is the first cloud-based digital laboratory informatics system to offer radiofrequency labels that automatically register open dates and calculate expiry dates, digitally capturing lab reagent data in real-time. It can be easily set up and rapidly integrated into existing experimental workflows, and allows for simple reagent identity checks. Additionally, the system provides automatic alerts to help scientists avoid using expired reagents and minimising experimental error.

BSSN had previously been involved with the AnIML standard. Will you continue to pursue standards for laboratory data?
Merck is absolutely committed to the continued development and support of an open and interoperable ecosystem built on AnIML and standardisation in lab automation. BSSN Software, in particular, is a key player in this space, being the first company to deliver a complete solution of tools for vendor-neutral management of analytical and biological data based on AnIML, and we are actively partnering with organisations and instrument manufacturers seeking to implement this framework.

If our current climate is an indication, it is more important than ever that data is accessible and that scientists are able to collaborate and share necessary data, either with their own internal colleagues or with external collaborators. Our BSSN Software solution now includes a new cloud-based analytical data viewer, which will make these kinds of virtual collaborations significantly easier.

Managing biological data

With laboratories creating ever-increasing amounts of data, it is critical to employ the correct strategy to ensure effective data management, writes Robert Roe

Alexander Supertramp/Shutterstock.com

The challenges of deriving insights from research data are often created at the point of recording and storing data from initial experiments. If an organisation is creating data silos which contain disparate file formats and data types, then going back to this data and using it for secondary research becomes challenging. Recording data in this way can lead to inefficiency, as files will often be duplicated and versioning issues can arise.

Modern laboratory instruments and practices make it easier than ever to create huge amounts of data. The challenge is creating 'good data' and storing it, so that it can be used in the most efficient way possible. To overcome this, laboratories must adopt some kind of data management strategy based on selecting the right software tools and creating a centralised repository for their experiments and analysis.

Many laboratories will have some kind of strategy for how to best record and store data effectively, but may not keep pace with the rise in data created by an organisation. Alternatively, they may not take into account how the data might be used in the future. At one time it was perfectly reasonable to save lots of experimental data on spreadsheets, but these practices have become untenable.

Maximising the effectiveness of research data

In 2019, Jisc, a UK provider of shared digital infrastructure and advice on digital technology for education and research, published a report highlighting the need for researchers to properly collect and manage data.

Report author Caroline Ingram, product lead at Jisc, said: 'Effective data management is carried out for the entire lifecycle of the data, from the point of creation through to dissemination, publication and archiving. Aspects of data management will usually continue long after the initial research project has ended.'

This can provide benefits to collaboration, research efficiency and research integrity through making data more accessible and easier to interpret.

> "Effective data management is carried out for the entire lifecycle of the data, from the point of creation through to dissemination, publication and archiving"

'Good research data management will enable you to organise your files and data for access and analysis without difficulty. This way you can track progress more easily, and mitigate against the risk of a team member leaving, taking valuable knowledge about the nature and extent of work completed with them.

'Good data management can result in improved research integrity, as well as act as validation for research results. Accurate and complete research data are an essential part of the evidence necessary for evaluating and validating research results, and for reconstructing

the events and processes leading to them.'

It is a complex challenge and many organisations are only just now realising the potential to use the tools that are available to them. Capturing data in a structured and standardised way in a centralised location, is what allows organisations to start deriving actionable insights from that data. There is little point in capturing all this data if you don't do that hard work to capture it in a way that is usable.

Laboratory information management systems (LIMS) and electronic laboratory notebooks (ELN) software can provide a good starting point for data management, but this depends on the workloads and applications that are being used in a specific laboratory.

In some cases, LIMS and ELN software will require additional integration with other software tools for analysis, library or registration of molecular data which can create additional steps in ensuring data is consistent.

Consolidating and bringing data into a centralised LIMS system can help improve data quality and integrity but is likely to have inconsistencies and duplicates, which requires manual data cleansing.

For scientists working with pharmaceutical or biotechnology data such as sequencing results, fermentation results, omics data and other associated metadata, this challenge can be particularly acute. For laboratories engaged in these disciplines, a single platform that can handle experimental results, associated metadata, sequencing and design tools could be the answer.

While legacy software is built according to outdated, potentially inaccurate notions of how scientists and organisations work, modern software tools for laboratory scientists need to reflect how scientists work today.

Benchling provides laboratory notebook alongside analysis tools for molecular biology, inventory management and registry functionality into a single platform to centralise biotechnology data.

This enables scientists to access data either in cloud-based software tools for digital DNA sequence editing, designing and running experiments, analysing data, and sharing research.

James Broughton, research lead at Mammoth Biosciences, said: 'Our Crispr-enabled platform for disease detection and bio-sensing grapples with the tremendous complexity of biological systems. Thanks to the scalability of the cloud, we can now explore massive amounts of DNA data and foster important innovations that promise to transform a range of industries – from healthcare to agriculture and beyond.'

Arsenal Bio's CSO and scientific co-founder Dr Theo Roth, said: 'At ArsenalBio, we're building the next generation of programmable cell therapies to bring these curative treatments to many more patients.

'This involves extensive novel DNA design and construction work, and Benchling has created ideal tools for our teams to collaboratively complete these complex molecular projects.

'It has accelerated the development of our gene-editing technologies and the synthetic DNA sequences that will power the next generation of cellular therapies,' he said.

For modern biotechnology workflows, controlling and managing the creation of data so that it can be effectively reused and shared is crucial to maintaining productivity and increasing efficiency. Utilising modern laboratory tools can expedite this process, enabling organisations to get ahead of the curve and manage data effectively.

Changing CFD

GEMMA CHURCH INVESTIGATES HOW CFD IS PROVIDERS ARE LOWERING THE BARRIER FOR SIMULATION

Computational Fluid Dynamics (CFD) is a staple fluid mechanics technology, where numerical analysis and data structures are used to analyse and solve a host of fluid flow problems. This established solution is now branching out, tackling everything from environmental issues to heart conditions and aeroacoustics.

This diversity of applications has affected the design and functionality of traditional CFD software, where engineers can now take on diverse problem sets with simulation and modelling. Andy Fine, Altair's vice president for CFD solutions, EMEA, said: 'We are dealing with increasingly complex physics when looking at CFD problems. While most CFD technology is extremely mature, it's very expensive and requires a lot of specialist knowledge. We are trying to lower these barriers to entry by providing increasingly accessible, cheaper solutions.'

The accuracy versus performance trade-off is another major challenge to democratise CFD, where some applications require large computational resources to meet user expectations. Ed Fontes, CTO at Comsol, explained: 'Here, you often find a conflict between the user's expectation and the software's ability to deliver accurate results to a small computational cost. Unfortunately, there are no free lunches: Accuracy costs.'

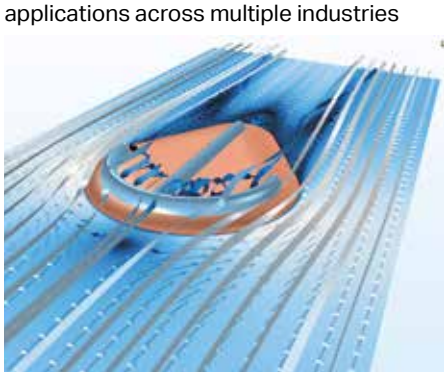
Fontes added: 'It is easy to get colourful plots at low computational costs by adding lots of artificial diffusion, but this yields inaccurate results. We invest a substantial portion of our development time in numerical methods and solvers, in order to solve problems accurately and on desktop computers.'

Multiphysics coupling is another increasingly important consideration in CFD, where tools must address usability issues. Fontes explained: 'It should be straightforward to couple CFD to other types of analysis, for example structural mechanics (fluid-structure interaction), chemical reactions, heat transfer, electromagnetic fields, equation-based modeling, etc. The solver settings for CFD should work together with the solution for other numerical equations added by other phenomena, either fully coupled or using a segregated solution strategy.'

Comsol's Application Builder is addressing these usability concerns. Fontes said: 'It allows CFD experts to build dedicated graphical user interfaces on top of specific CFD problems defined in Comsol. These experts can build applications that can be used by a larger community of engineers and scientists that are domain experts, but may not be CFD experts.'

Comsol has also optimised many of its existing tools for CFD over the last few years, including its meshing tool, iterative solvers and tools for results evaluation. 'The mesh and solver projects have been very important for CFD and are still large projects, so we are in no way done. We are proud of the ability to visualise and evaluate results,' Fontes explained.

Altair has recently increased its focus on CFD, targeting specific and specialist applications across multiple industries using a suite of software solutions. Fine explained: 'One size does not fit all in the CFD space. We make a decision for each client, based on their application, to find the best solution.'



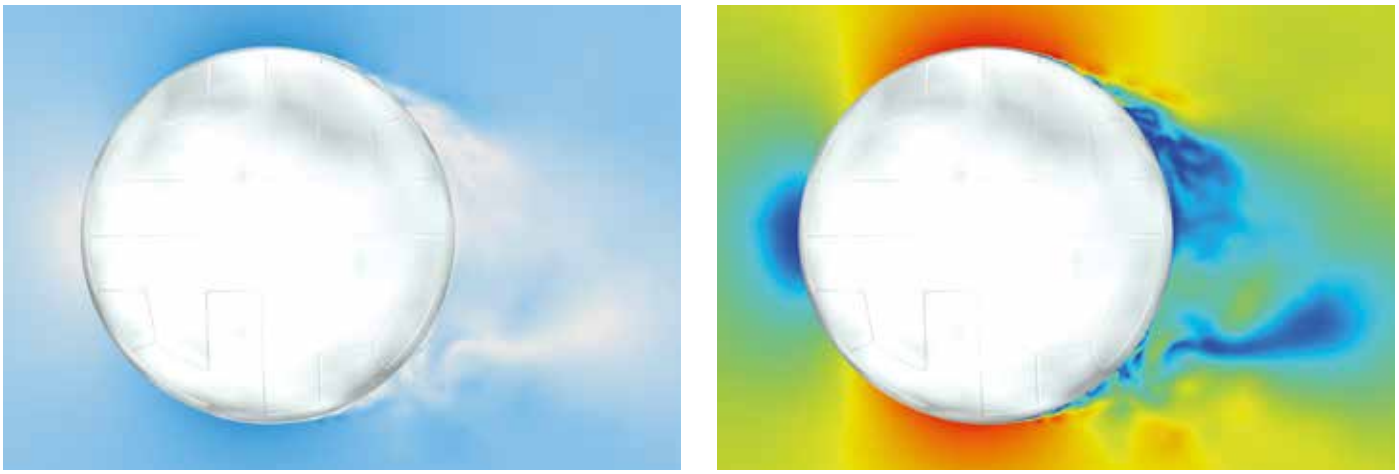
Model of the CFD around a prehistoric animal. The optimal position of the feeding organ (sort of mouth) would be in the recirculation zone behind the arms of the anchor-shaped structure, where solid nutrients would settle on the surface of the animal. (Model reproduced from Darroch et al, https://royalsocietypublishing.org/doi/full/10.1098/rsbl.2017.0033)

Mixing and free surface shape in a chemical reactor. The impeller is placed at the bottom and two heat exchanger tubes are placed around the impeller. The model can be used to compute the shape of the free surface, flow field, composition, and temperature field in the reactor

Altair runs a unique licensing system called Altair Units. The company's full suite of software solutions is available via Altair Units from its marketplace Altair One, where engineers can also get access to specialist support teams, including a CFD division, and third-party solutions from the Altair Partner Alliance. Altair's CFD simulation suite includes Altair AcuSolve, which allows engineers to tackle several general purpose CFD problems, including flow, heat transfer, turbulence, and non-Newtonian material analysis, for example.

Altair ultraFluidX and Altair nanoFluidX are application-specific CFD solutions. Altair ultraFluidX was developed specifically for ultra-fast prediction of aerodynamic properties for a variety of vehicles, whereas Altair nanoFluidX is a particle-based hydrodynamics (SPH) tool to predict fluid flow around complex geometries under complicated motion. 'These tools allow us to tackle problems that are practically impossible to work with using traditional CFD,' Fine said.

One of these emerging areas is aeroacoustics, where the noises and vibrations created by an object in motion are investigated and reduced. These sounds are now more noticeable than



Flow field around a football travelling at 20m/s (the speed of a cross that you can head). This looks like a relatively simple problem but it requires millions of degrees of freedom and advanced models, in this case large eddy simulations (LES), in order to obtain relatively accurate values of drag and lift. The solution shown in this figure takes days to compute on a desktop computer just to simulate half a second of flight

ever, thanks to the development of electric vehicles, which provide a much quieter ride, compared to traditional vehicles.

The field of aeroacoustics is also increasingly relevant in other industries, including machining and tooling where manufacturers want to reduce noise created by large cooling fans, for example. Fine said: 'This is a key and unique application area where we have the domain knowledge and understanding to solve these issues, not just simulate them, but also predict corrections and geo-spatial modifications that can solve any number of problems.'

The fields of plasma science and electrochemistry have both expanded significantly in recent years, according to Fontes, who explained: 'Here, we find problems involving CFD, electromagnetic fields, chemical, and electrochemical reactions, but also the usual conjugate heat transfer problems (coupled heat transfer in solids and fluids with CFD in the fluids) that emerge in all processes where heat is required or generated.'

Biomedical applications such as heart pumps and other medical implants are also 'an interesting area of application,' according to Fontes, who added: 'Medtech is a fascinating new area for CFD. If we look at the traditional fields where simulation and modelling is often used, they are simple systems, compared to biological systems. Now, thanks to advances in CFD technology, we can deal with increasingly complex biological systems.'

Comsol Multiphysics was recently used to simulate the HeartMate 3 left ventricular assist device (LVAD) from Abbott Laboratories. In patients with a poorly functioning left ventricle, an LVAD pump is responsible for circulating oxygen-rich blood through the body. CFD and electromagnetics modelling helped the Abbott team design a powerful, efficient and hemocompatible pump with pulsatile flow, which more closely mimics a functioning heart.

Freddy Hansen, a physicist at Abbott Laboratories, said: 'It is easy to enter your own mathematical expressions in Comsol, in particular as you can type in an equation anywhere you can enter a number, and for us this greatly simplifies comparing different designs for their hemocompatibility. To do this, we enter custom algebraic and differential equations and solve these alongside the CFD. We find that the relatively large gaps that we have in the HeartMate 3 between its rotor and the surrounding walls – gaps that are only possible with a magnetically levitated rotor – offer the design that is the most gentle on the blood.'

> "Thanks to advances in CFD technology, we can deal with increasingly complex biological systems"

Such work could revolutionise the field of medicine, as Fontes concluded: 'Medicine today is very much trial and error but, using CFD, we could analyse a model and then create treatments that are not only cheaper, but adapted to an individual's symptoms and physical characteristics.'

Environmental technology is another growing CFD field 'involving processes such as carbon sequestration and processes related to the hydrogen society,' according to Fontes. 'CFD and fluid-structure interaction of wind turbines has been a popular application for at least 10 years, but also the design around the wind turbines becomes interesting. For example, acoustics, water flow for offshore farms, and airflow around wind farms presents many interesting CFD and multiphysics problems involving, for example, fluid-structure interaction.'

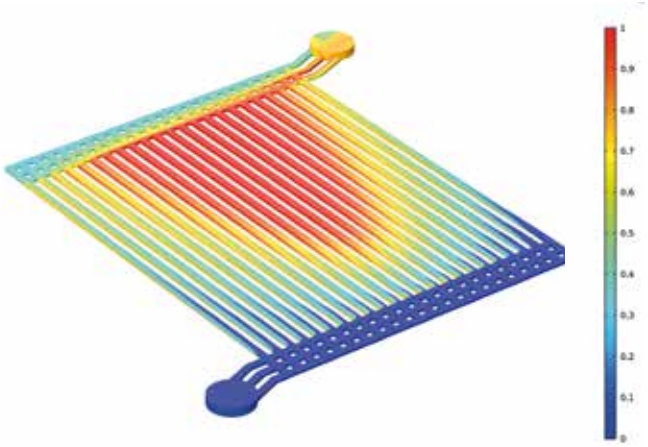
Smokin' CFD

The Barcelona Supercomputing Center (BSC) is currently working on the Estimate project, a European initiative from the Clean Sky Joint Undertaking. Researchers are developing advanced CFD software to help the design of the new generation of efficient and sustainable Ultra High Bypass Ratio engines, with a focus on soot emissions.

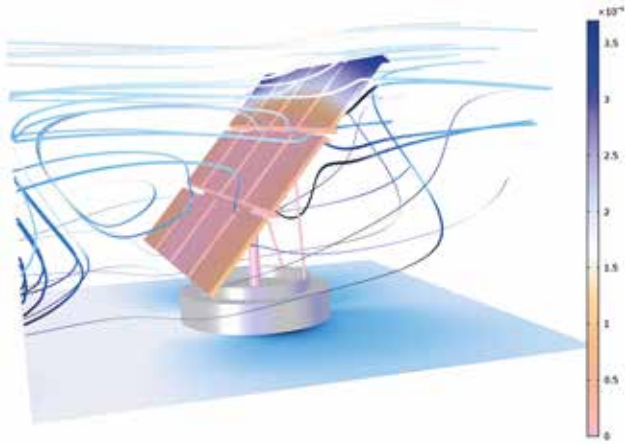
When at full power conditions, soot emissions tend to be low but, at intermediate conditions, particulate formation can be significant. Estimate uses a multidisciplinary approach to predict soot formation from chemistry oxidation to particle formation, based on high-fidelity simulations validated with reference experiments.

The project will shed light onto the influence of the combustion model and the treatment of turbulence chemistry interactions on the prediction capabilities of the soot models, in terms of soot volume fraction, soot volume density and particle size distributions.

The team is currently developing these models, and is addressing a range of challenges in its work. Oriol Lehmkuhl, large-scale computational fluid dynamics group leader at BSC, said: 'We do not have a good model to predict combustion and turbulence together with soot formation and evolution. We have challenges on the chemical, modelling and computational →



Oxygen gas volume fraction in the anode channels in a polymer electrolyte membrane electrolyzer. The model is used to compute the two-phase flow field in the bipolar plate channels



Fluid-structure interaction on the surface of a solar panel unit. The wind forces the displacement of the panels, and the risk for failure due to fatigue can be estimated using a coupled CFD and structural mechanics analysis

→ side because we need to compile a lot of different models while maintaining efficiency in HPC facilities.'

Lehmkuhl added: 'We also have an extra challenge where, usually soot formation modelling is done with gas [engines] but, with kerosene, we have to take into account multi-phase conditions.'

There is also an industry-led requirement for this work, as Daniel Mira, propulsion technologies group leader at BSC, explained: 'The aerospace industry has clear targets for CO_2 emissions but now the ones for soot emissions are becoming increasingly important. The

> "We do not have a good model to predict combustion and turbulence together with soot formation and evolution"

prediction of particle size distribution is a major challenge in the aeroengine industry, where standard prediction tools have massive errors between 200 and 300 per cent. We would like to see how these advanced soot models can predict soot formation at aeroengine-relevant conditions'

The team is making good progress but there are also many experimental considerations to take into account, where the team will eventually need to validate the model with experiments that successfully mimic real-world conditions. Lehmkuhl explained: 'We must be able to correctly measure the soot formation and distribution in both laminar and turbulent flames, and this is very complex.'

The validation and assessment of the models will come from the definition of dedicated experiments with an increased level of complexity, going from pre-vaporised counterflow flames, kerosene spray flames up to single sector and full annular rigs.

The project includes the development of reaction mechanisms for the fuel oxidation and soot precursors, and the development of a phenomenological model for primary breakup of airblast atomisers. 'We have a lot of different physics that we need to put together, and we need to validate as much as is possible,' Lehmkuhl added.

AI Potential

Could CFD benefit from an AI-enabled upgrade? Polish software specialist byteLake is developing algorithms to optimise existing CFD workflows and lower the cost of such solutions.

The byteLake CFD Suite features a collection of AI models, some replace traditional numerical solvers with equivalent AI models, while others accelerate CFD solvers with AI algorithms. This software suite promises to expedite the analysis process dramatically, providing 'immediate results', according to Marcin Rojek, co-founder at byteLake, who added: 'Artificial intelligence is the next step for CFD, unlocking fast simulations.'

The firm is currently developing its CFD suite using steady-state simulations, before moving to transient simulations. 'For steady-state, we have not seen any limitations where AI cannot be used,' Rojek added.

The company recently completed its first accuracy tests with 24 different simulations using OpenFoam. A steady-state icoFoam simulation was carried out with a mesh size of 400 cells. The predicted results achieved a mean squared error of less than 0.001 and a Pearson correlation coefficient of more than 0.95, with more than 99 per cent relative accuracy. byteLAKE is now investigating larger mesh sizes and accelerating other OpenFoam solvers.

Mariusz Kolanko, co-founder at byteLake, said: 'The key challenge is to accelerate CFD simulations beyond what is currently possible. We are also working to ensure the AI models are easy to use and compatible with existing workflows.'

The CFD Suite is compatible with OpenFoam and uses TensorFlow-compatible algorithms to boost its interoperability, where TensorFlow is a leading open source machine learning platform. byteLAKE is working towards providing a plug-and-play interface, where CFD engineers should not need to change data types, replace toolchains, etc, while providing cross-platform compatibility, with no hardware or infrastructure upgrades required to run the software.

Ease-of-use is another important deliverable for the CFD Suite, where engineers should just prepare their input data in the same way that they would for a traditional CFD solver but, instead, send it across to an AI-enhanced CFD model.

'We want to make a tool that will be compatible with all the major CFD players and is also easy to use,' Rojek said. The CFD Suite is scheduled for release in November 2020.

While challenges remain for today's CFD solutions in terms of usability, validation and performance for a given accuracy, the inclusion of AI technologies is an interesting, new direction. This specialist work and other innovations are now more important than ever to address an increasing range of real-world problems in the CFD space.

White Papers
now available online

VIEW FOR FREE*

The Role of a Cloud LIMS in the Digital Transformation of the Laboratory
By Interactive Software Limited

Implementing a cloud LIMS isn't a mandatory step in becoming 'digital'. However, its underlying infrastructure and technology can jump start and be an integral component of your digital transformation.

Data Integrity: Technical Controls That Demonstrate Trust
By Thermo Fisher Scientific

Data Integrity: Technical Controls That Demonstrate Trust. Achieve compliance and maintain data integrity with the latest Thermo Scientific™ Chromeleon™ 7.3 Chromatography Data System (CDS),which provides all the necessary preventative and technical controls, built into one

AI: 5 Steps to get started
By Nvidia

Looking to get started with AI but not sure where to begin? Read NVIDIA's quick start guide and learn how to set your next AI project up for success.

Application porting and code-optimization activities for European HPC systems
By Prace

Currently, the HPC community is actively working towards the development of next-generation HPC systems capable of delivering ExaFLOP performance. While these new HPC architectures open exciting opportunities to the computational scientific community, they also pose numerous challenges.

How can biotech companies optimize data access while going virtual?
By Collaborative Drug Discovery (CDD)

There is mounting pressure for companies to ensure that non-essential personnel work from home.

Choosing a LIMS or SAP for your Laboratory
By Abbott Informatics

Would a LIMS or SAP better meet the needs of your company? In this whitepaper, we will take a look at the different systems on the market for managing laboratory operations.

*Registration required

www.scientific-computing.com/white-papers

SCIENTIFIC COMPUTING WORLD

Suppliers' Directory

Find the suppliers you need quickly and easily
www.scientific-computing.com/suppliers

HIGH-PERFORMANCE COMPUTING

Altair Engineering, Inc
(248) 614-2400
info@altair.com
www.altair.com
■ Software

Boston Limited
+44 (0)1727 876 100
sales@boston.co.uk
www.boston.co.uk
■ Systems Integrator

Calyos SA
0032 4 75 73 54 65
newbusiness@calyos-tm.com
www.calyos-tm.com
■ Cooling

Cluster Vision
+31 20 407 7550
marketing@clustervision.com
www.clustervision.com
■ Software

Eurotech Spa
+39 0433 485 411
welcome@eurotech.com
www.eurotech.com
■ Systems Integrator

Numerical Algorithms Group (NAG)
+44 (0)1865 511 245
katie.ohare@nag.co.uk
www.nag.co.uk
■ Software

OCF
+44 (0) 114 257 2200
scw@ocf.co.uk
www.ocf.co.uk
■ Systems Integrator

LABORATORY INFORMATICS

Abbott Informatics
+44 (0)161 711 0340
AISalesUK@abbott.com
www.abbottinformatics.com
■ LIMS
■ Scientific Document Management Systems

Accelerated Technology Laboratories
+1 910 673 8165 (outside of US)
800 565 LIMS (5467) US and Canada
info@atlab.com
www.atlab.com
■ Bioinformatics
■ LIMS

Amphora Research Systems
+44 (0)845 230 0160
info@amphora-research.com
www.amphora-research.com
■ Electronic Laboratory Notebooks

Autoscribe
+44 (0) 118 984 0610
info@autoscribeinformatics.com
www.autoscribeinformatics.com
■ Data Management
■ Electronic Laboratory Notebooks
■ LIMS
■ Scientific Document Management Systems

Biovia
+ 1 858 799 5000
www.3ds.com/products-services/biovia
■ LIMS
■ Electronic Laboratory Notebooks

Eusoft
+39 080 5426799
info@eusoft.it
www.eusoft.it
■ LIMS

IDBS
info@idbs.com
www.idbs.com
■ Electronic Laboratory Notebooks

InfoChem
+49 89 2153 7160
info@infochem.de
www.infochem.de
■ Cheminformatics

LabWare
infoEU@labware.com
www.labware.com
■ LIMS

Osthus GmbH
+49 241 943 140
office@osthus.de
www.osthus.de
■ Electronic Laboratory Notebooks
■ Cheminformatics
■ LIMS

Siemens
+1 322 536 2139
industrial-it.swe@siemens.com
www.siemens.com/simaticit-rdsuite
www.siemens.com/industrial-it/lims
■ Electronic Laboratory Notebooks
■ LIMS

Thermo Fisher Scientific
+44 (0)161 942 3000
marketing.informatics@thermofisher.com
www.thermoscientific.com/informatics
■ LIMS
■ Scientific Document Management Systems
■ Chromatography Data Systems
■ Spectroscopy Software

MODELLING AND SIMULATION

Altair Engineering, Inc
(248) 614-2400
info@altair.com
www.altair.com
■ Software

Integrated Engineering Software
+1 204 632 5636
info@integratedsoft.com
www.integratedsoft.com
■ Mathematics, Simulation and Modelling

FREE ON-DEMAND WEBCAST*
AVAILABLE NOW

Enabling Discovery In Research With Cloud Technologies

SPEAKERS

Prof. Imre Berger
Director, Max Planck Centre for Minimal Biology, University of Bristol, UK

Dr Dan Ruderman
Director of Analytics and Machine Learning, at Lawrence J. Ellison Institute for Transformative Medicine of USC, USA

Eric Grancher
Head of database services group, CERN IT department, at CERN, Switzerland

Taylor Newill
Senior Manager, HPC Product Management, at Oracle

Research drives the global expansion of knowledge and spurs scientific, technological and social innovation. Currently, all too often research investigations are constrained by the limitations of only using traditionally available computing technologies. Existing technologies may not be the most advanced, they may have capacity thresholds and typically will have a queuing system in place. These conditions impede progress and impact research outcomes. Cloud technologies, by contrast, are now opening up entirely new possibilities, enabling researchers to access additional, powerful resources and alternative technologies, exactly as and when they are required. Researchers today thus have many more options to help them accelerate time to discovery and advance innovation.

Listen to fascinating insights about scientists and research organisations using, or planning to use, cloud technologies, to advance their ground-breaking research endeavours.

SCIENTIFIC COMPUTING WORLD

ORACLE

www.scientific-computing.com/webcasts

*This webcast is FREE to view, though registration is required

LABWARE

LIMS and ELN in a single platform
Automation software for all laboratories

LABWARE
Results Count

Offices worldwide supporting customers
in more than 125 countries

www.labware.com