28-29/3/2019

Title: Container Technologies in Cloud and High Performance Computing Research and Commercial Applications
Location: Eulerzaal - Amsterdam Science Park Conference centre; Science Park 125, 1098 XG Amsterdam (NL)

Thursday, 28 March

Start	End	Duration	Description	Affiliation
12:20:00 PM	12:30:00 PM	0:10	**Welcome and introduction**	
12:30:00 PM	12:55:00 PM	0:25	**Peter Coveney**	University College London
12:55:00 PM	1:20:00 PM	0:25	**Mariano Vasquez**	Barcelona Supercomputing Center
1:20:00 PM	1:45:00 PM	0:25	**Alexandre Bonvin**	Utrecht University
1:45:00 PM	2:10:00 PM	0:25	**Reginald Cushing**	University of Amsterdam
2:10:00 PM	2:25:00 PM	0:15	**Coffee break**	
2:25:00 PM	2:50:00 PM	0:25	**Andrew Lahiff**	UKAEA
2:50:00 PM	3:15:00 PM	0:25	**Phil Bates**	Oracle
3:15:00 PM	3:40:00 PM	0:25	**Jorge Gomes**	LIP
3:40:00 PM	3:55:00 PM	0:15	**Coffee break**	
3:55:00 PM	4:20:00 PM	0:25	**David Godlove**	Sylabs Inc.
4:20:00 PM	4:45:00 PM	0:25	**Timothy Randles**	Los Alamos National Lab
4:45:00 PM	5:10:00 PM	0:25	**CJ Newburn**	NVIDIA
5:10:00 PM	5:40:00 PM	0:30	**Panel discussion**	
			Social dinner	

Friday, 29 March

Start	End	Duration	Description	Affiliation
9:00:00 AM	9:25:00 AM	0:25	**Christian Kniep**	QNIB Solution
9:25:00 AM	9:50:00 AM	0:25	**Marian Bubak**	AGH University of Science and Technology
9:50:00 AM	10:15:00 AM	0:25	**Björn Backeberg**	EGI
10:15:00 AM	10:40:00 AM	0:25	**Coffee break**	
10:40:00 AM	11:05:00 AM	0:25	**Kenneth Hoste**	Ghent University
11:05:00 AM	11:30:00 AM	0:25	**Alexandar Mechev, Frits Sweijen**	Leiden University
11:30:00 AM	11:55:00 AM	0:25	**Hamid Arabnajed**	Bunel University
11:55:00 AM	12:55:00 PM	1:00	**Lunch**	
12:55:00 PM	1:20:00 PM	0:25	**Joao Damas**	Acellera Ltd.
1:20:00 PM	1:45:00 PM	0:25	**Chris Reynolds**	Diamond Light Source
1:45:00 PM	2:10:00 PM	0:25	**Charles Laughton**	University of Nottingham
2:10:00 PM	2:30:00 PM	0:20	**Coffee break**	
2:30:00 PM	3:00:00 PM	0:30	**Panel discussion**	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 675451 (CompBioMed project).

List of Abstracts

Presenter
Björn Backeberg (EGI)

Title
EGI Cloud Container Compute to support biomedical research

Abstract
The EGI Foundation aims to empower researchers from all disciplines to collaborate and carry out data- and compute-intensive science and innovation. To do this, EGI facilitates access to a federated Infrastructure as a Service (IaaS) cloud for researchers to run compute- and / or data-intensive tasks and host online services in virtual machines and / or Docker containers. EGI brings together the largest community cloud federation in Europe with 20 cloud providers across 14 European countries offering IaaS computing and storage services. The EGI Cloud Federation aggregates resources by defining a set of federation features on top of the local IaaS interfaces and protocols – such as resource discovery, user authentication, monitoring, and VM image distribution – enabling workloads to span and seamlessly migrate across resource centers. Through the EGI Application Database – EGI offers the possibility to dynamically deploy cloud applications in a federated cloud infrastructure using a simple GUI. The EGI Cloud Container Compute service (https://www.egi.eu/services/cloud-container/) builds on the EGI Cloud Compute service to give users the ability to deploy and scale Docker containers on-demand using Kubernetes. It offers computational resources in a secure and isolated environment with industry-standard API access, without the overhead of managing the operating system. During this presentation we will give an overview of the EGI federated cloud, provide details about at the EGI Cloud Container service and the technologies it is built upon, demonstrating how to easily start Kubernetes clusters on EGI resources and present examples of community use cases.

Presenter
Phil Bates (Oracle Cloud)

Title
Cloud for Research & Higher Education

Abstract
High performance cloud computing is enabling new approaches for researchers to dramatically accelerate research. Phil will explain the key concepts and capabilities that researchers are using from high performance public cloud infrastructure and explain Oracle's programme which supports researchers and educators globally.

Presenter
Alexandre Bonvin (Utrecht Uni)

Title
Structural biology in the clouds: Automation through containerisation

Abstract
Structural biology deals with the characterization of the structural and dynamic properties of biological macromolecules and adducts thereof. Gaining insight into 3D structures of biomolecules is highly relevant with numerous applications in health and food sciences.
Since 2010, the WeNMR project (www.wenmr.eu) has implemented numerous web-based services to facilitate the use of advanced computational tools by researchers in the field, using the grid computational infrastructure provided by EGI. These services have been further developed in subsequent initiatives under the H2020 EGI-ENGAGE, INIDCO-Datacloud and West-Life projects. The WeNMR services are currently operating under the European Open Science Cloud with the H2020 EOSC-Hub project (www.eosc-portal.eu).
In my talk, I will illustrate the use of containers to both automate the deployment of our services and run complex software on HTC resources.

Presenter
Peter Coveney (UCL)

Title
High throughput binding affinity predictions enabled through heterogeneous workflows at the emerging exascale

Abstract
DOE INCITE The INtegrated and Scalable PredictIon of REsistanace (INSPIRE) project aims to lay the foundations for the use of molecular simulation and machine learning to guide precision cancer therapy, in which therapy is tailored to provide maximum benefit to individual patients based on the genetic information about their particular cancer. The core workflow combines elements with highly varied computational demands, including docking, molecular dynamics and active learning.

The project targets Summit, the world's most powerful supercomputer, which uses the novel Power Little Endian (POWER9 or ppc64le) architecture. This presents two distinct challenges; firstly the Summit design is based around large monolithic job execution and, secondly, not all of the components of the workflow have been ported to this architecture (and in some cases are unlikely to be in the near future). In order to allow us to maintain the freedom to use best in class software and to orchestrate our workflows we make use of the OpenShift container

application platform which is connected to the same network and storage infrastructure as Summit. We believe this combination of systems is an ideal platform for hybrid high throughput/high performance workflows which are likely to be a growing class of applications as we move towards the exascale.

Presenter Joao Damas (Accellera)

Title Containerization for Reproducible Deployment of Biomedical Applications and Workflows in Diverse Computing Infrastructures

Abstract
In recent years we have seen containerization technologies come of age. The interest in these technologies is multi-factored but on this talk we will focus on reproducibility and deployment of biomedical applications and workflows. The access to state-of-the-art methods by both research and application scientists in an easy and painless manner is crucial for their success and widespread use. Acellera has worked in providing those methods through applications and workflows on its PlayMolecule platform, where containerization technologies have been proved crucial. The importance of containerization in this context includes isolation of conflicting version and dependency trees, distribution of code that will work out-of-the-box in multiple types of computing infrastructures (from local workstation to HPC), and the reproducibility of results (at least deterministic code) in different infrastructures and different points in time.

Presenter Dave Godlove (Sylabs Inc.)

Title Singularity: simple, secure containers for HPC

Abstract
Singularity is the most widely used container solution in high-performance computing (HPC). Instead of a layered filesystem, a Singularity container is stored in a single file. This simplifies the container management lifecycle and facilitates features such as image signing and verification to produce trusted containers. At runtime, Singularity blurs the lines between the container and the host system allowing users to read and write persistent data and leverage hardware like GPUs and Infiniband with ease. The Singularity security model is also unique among container solutions. Users build containers on resources they control or using a service like the Sylabs Remote Builder. Then they move their containers to a production environment where they may or may not have administrative access and the Linux kernel enforces privileges as it does with any other application. These features make Singularity a simple, secure container solution perfect for HPC workloads. Sylabs Inc maintains Singularity and fosters the open-source community. Sylabs also offers a professionally curated and supported version of Singularity called SingularityPRO with enhanced security and stability for production-grade centers.

Presenter Jorge Gomes (LIP)

Title Running Linux containers with udocker

Abstract
udocker is a tool that addresses the execution of Linux containers in user space, i.e. without installing additional system software, without requiring any administrative privileges and in a way that respects resource usage policies, accounting and process controls. udocker aims to empower users to execute applications encapsulated in containers across a wide range of systems.

Presenter Christian Kniep (QNIB Solution)

Title Containerized Convergence of Big Data and Big Compute

Abstract
Biomedical disciplines relying on fast iterating, productivity driven workloads - always on the cutting edge of the IT community. The advent of GPGPU acceleration made computation (mostly AI/ML) economically feasible. Container technology provided the packaging and execution environment to lift this off the ground; allowing for reproducibility, collaboration and application and data lifecycle management.
The challenges faced and the technologies used are similar to what is discussed in the High Performance Computing (HPC) space and hence this talk will explain the convergence of Big Data and Big Compute by virtue of Hyperscale technologies.
First Christian will refresh the audience memory on what containerization is about, segueing into why AI/ML workloads are leading to fully fledged HPC applications eventually anyway and how this will inform the way forward.
Christian will discuss the first of the three challenges (`Hardware Access`, `Data Access` and `Distributed Computing`) in container technology and how it can be tackled by the power of open source.

Presenter	Title
Andrew Lahiff (UKAEA)	Running HTC & HPC applications opportunistically across private, academic and public clouds

Abstract
Access to both HTC and HPC facilities is vitally important to the fusion community, not only for plasma modelling but also for advanced engineering and design, materials research, rendering, uncertainty quantification and advanced data analytics for engineering operations. The computing requirements are expected to increase as the community prepares for ITER, the next generation facility. Moving to a decentralised computing model is vital for future ITER analysis where no single site will have sufficient resource to run all necessary workflows. PROMINENCE is one of the Science Demonstrators in the European Open Science Cloud for Research Pilot Project and aims to demonstrate that the fusion community can make use of distributed cloud resources. Here we will describe our platform which enables users to run jobs in containers opportunistically across a wide variety of cloud sites. Hierarchical cloud bursting allows jobs to burst from local private clouds onto national research clouds, to EGI/EOSC resources and finally through to public clouds, completely transparently to the user.

Presenter	Title
Charles Laughton (Uni Nottingham)	Crossbow: development of a platform and workflow system for cloud-based computational chemistry

Abstract
I will describe the current status of Crossbow, a Python-based application designed to provide biomolecular simulation scientists with 'one-click' deployment of bespoke compute clusters in the cloud, and the development of Crossflow, a workflow system tailored to this environment. Containerization is playing a major role as we seek to balance ease of deployment, configurability, and the maintenance burden.

Presenter	Title
Alexandar Mechev, Frits Sweijen (Leiden Uni)	Streamlined deployment of astronomical software

Abstract
Building, testing and deploying complex software on a distributed system is difficult. Several steps are necessary for verification and validation of the software. This includes building and testing software packages, integration tests for scientific pipelines and verification of data quality. Performing these steps is time consuming for data processing pipelines that change daily. The result is a set of ever-changing software packages that produce results of unknown quality. We present a method to build and test astronomical software on a distributed environment using version control (git) and containerization (singularity). Our system can make daily builds of our software, perform unit and integration tests and version the resulting images. The resulting images will make astronomical processing reproducible.

Presenter	Title
Timothy Randles (LANL)	Charliecloud: Los Alamos National Laboratory's Lightweight Unprivileged Container Runtime

Abstract
Supercomputing centers are seeing increasing demand for user-defined software stacks (UDSS), instead of or in addition to the stack provided by the center. Linux container technology is the most promising way to achieve the portability and consistency required by users who want to provide their own UDSS. We will discuss common myths and misconceptions about Linux containers before presenting Charliecloud. Charliecloud uses the Linux user and mount namespaces to run industry-standard Docker containers with no privileged operations or daemons on center resources. Our simple approach avoids most security risks while maintaining access to the performance and functionality already on offer, doing so in just 800 lines of code.

Presenter	Title
Chris Reynold (Diamond)	Containerization at Diamond

Abstract

Diamond Light Source is the UK's national synchrotron. This talk will cover the current and future use of containerization technologies at Diamond. Use cases for containers at Diamond are broad ranging. They include monitoring systems, traditional HPC used to support the scientific workflow, and dev-ops pipelines to enable rapid development of Diamond scientific software.