

## Case Study: EPCC and StackHPC deliver a highly-available data science cloud

The Edinburgh International Data Facility (EIDF) is recognised as a state-of-the-art, high-powered data analytics and storage service. EPCC built and manages EIDF specifically to underpin the Data Driven Innovation (DDI) initiative as part of the Edinburgh and South-East Scotland City Region Deal. To support a wide array of research and innovation projects, EPCC sought a robust, flexible private cloud environment.



StackHPC was engaged by EPCC to provide a "from design to support" solution for the EIDF Data Science Cloud (DSC). This collaboration successfully delivered a resilient, highly targeted software-defined environment specifically tailored for the research community.

### The Challenge and Context

EPCC sought to establish the EIDF Data Science Cloud as a private, on-premise cloud infrastructure. This new environment needed to operate standalone while ensuring seamless integration with existing High-Performance Computing (HPC) services and storage platforms at EPCC.

The requirements for the DSC were substantial:

- **Scale and Scope:** The design needed to support approximately 200 user projects. Furthermore, larger collaborations could require up to 50 Virtual Machines (VMs) and support around 1,000 users.
- **Flexibility and Functionality:** The primary initial purpose was to provide VM services. Crucially, the cloud had to offer greater flexibility than traditional production HPC infrastructure, providing environments for users developing novel approaches or non-conventional software stacks. This included offering compute accelerators such as NVIDIA Graphics Processing Units and Cerebras' Wafer Scale Engine.
- **Infrastructure Management:** To ensure maintainability and consistency, infrastructure deployment and configuration had to be highly automated and defined as version-controlled source code.

## The Collaborative Solution: OpenStack Infrastructure and Automation

Working in collaboration with EPCC staff, StackHPC designed and deployed the core infrastructure for the Data Science Cloud. The solution is built upon **OpenStack Infrastructure as a Service (IaaS)**, which is widely adopted within Research Computing.

The deployment leveraged robust automation and high-availability features:

- **Deployment and Orchestration:** The entire deployment process is driven by Ansible. The OpenStack control plane deployment and configuration utilised OpenStack Kayobe, an Ansible-driven framework, to provide high-availability
- **Hardware Provisioning:** The deployment process is initiated from a Deployment Seed Node, which instructs Kayobe to configure and deploy the control plane nodes. The physical infrastructure includes three types of hypervisors: Mid-Range, High-End, and GPU hypervisors.
- **Monitoring and Telemetry:** Control plane monitoring is managed using Prometheus. Log aggregation is implemented using an EFK stack (Elasticsearch, Fluentd, and Kibana).

## Storage Integration with Ceph

Ceph storage forms a critical component of the infrastructure, used to back key OpenStack services. Ceph provides the backing storage for Glance (the image service) and Cinder (the volume storage service). User instances are also configured to access various Ceph block storage systems such as S3 and CephFS.

Furthermore, in a separate engagement, StackHPC assisted EIDF in migrating an existing Ceph cluster from a vendor appliance to a fully open-source solution, using similar Ansible orchestration as to the OpenStack infrastructure component.

## Integration with EPCC Systems and User Access

A critical measure of success was the integration of the new OpenStack environment with EPCC's established administrative and resource systems:

- **User Authentication (SAFE):** OpenStack API users authenticate using the SAFE federated service via OpenID Connect (OIDC). The EPCC SAFE system (Software framework for Advanced computing Facilities) handles resource management, accounting, reporting, and usage monitoring.
- **User Access Layer (EIDF Portal):** EPCC developed the EIDF Portal, a self-service platform that allows project managers to provision and destroy VMs, handle project user accounts and assigned storage within their assigned quotas.

## Primary Use Cases and Functionality

The EIDF Data Science Cloud focuses on supporting research and innovation where maximum flexibility is required:

- **Flexible VM Provisioning:** The primary initial purpose is the provision of VM services for researchers, industry partners, public sector organisations, and students. Researchers use the DSC to host and process their project's data. Also, EIDF provides users with EIDF Notebooks, which offer low-entry access to working with Python, R and Julia.
- **Novel Development:** The DSC is crucial for users seeking to develop novel approaches or utilise non-conventional software stacks, as this allows them to bypass the limitations of EPCC's production HPC infrastructure.
- **Specialised Compute:** The cloud provides accelerated compute resources via Kubernetes, enabling uses like PhD students testing their ideas on a GPU-accelerated node. Currently the EIDF offers several NVIDIA GPU models and the Cerebras CS3 Waverscale Engine.
- **Future Development:** Subsequent phases of the DSC are intended to include related functionality, such as support for secure research computing platforms (Secure Virtual Desktops) and inference support.

## Outcomes and Sustained Support

The successful collaboration resulted in the delivery of the OpenStack-based EIDF Data Science Cloud, enabling EPCC to support high-level research and data-driven innovation.

Following deployment, StackHPC provided training to EPCC staff to facilitate day-2 operations and promote greater autonomy for EIDF operations. StackHPC maintains an ongoing role, continuing to provide support and maintenance services for both the OpenStack and Ceph clusters, ensuring the stability and longevity of the EIDF infrastructure.