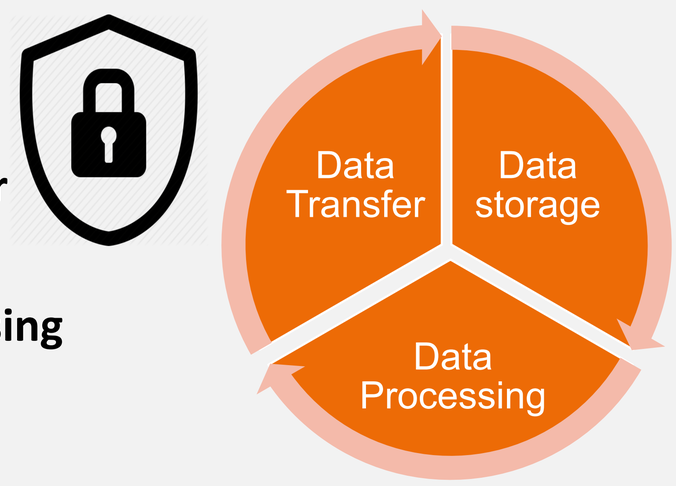
Secure Data Processing on shared HPC systems

High performance computing (HPC) clusters operating in shared and batch mode pose challenges for processing sensitive data. Our platform as a service solution provides a customisable virtualized solution that addresses this without modifying existing HPC infrastructures. Using PCOCC and SLURM this platform can be used for processing sensitive data within a shared HPC environment and address both strict and flexible data security requirements.

Background and motivation

Shared high-performance computing facilities are very popular due to their accessibility and cost-effectiveness. However, this makes it challenging to impose security requirements for processing highly sensitive data. With growing awareness of data security and privacy, there is a high demand for data processing, storage and transfer facilities with a tightly managed security level.

- ✓**Secure data transfer**
- ✓**Secure data storage**
- ✓**Secure data processing**

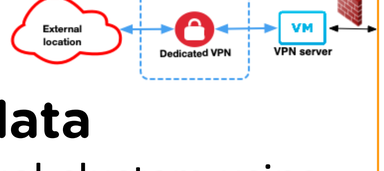


Why PCOCC?

- ✓Build a private cloud on your existing publicly accessible cluster
- ✓Build private virtual clusters with negligible performance penalty
- ✓Offer fully customized environment meeting functional and security requirements
- ✓Integration with widely used open source batch scheduler SLURM

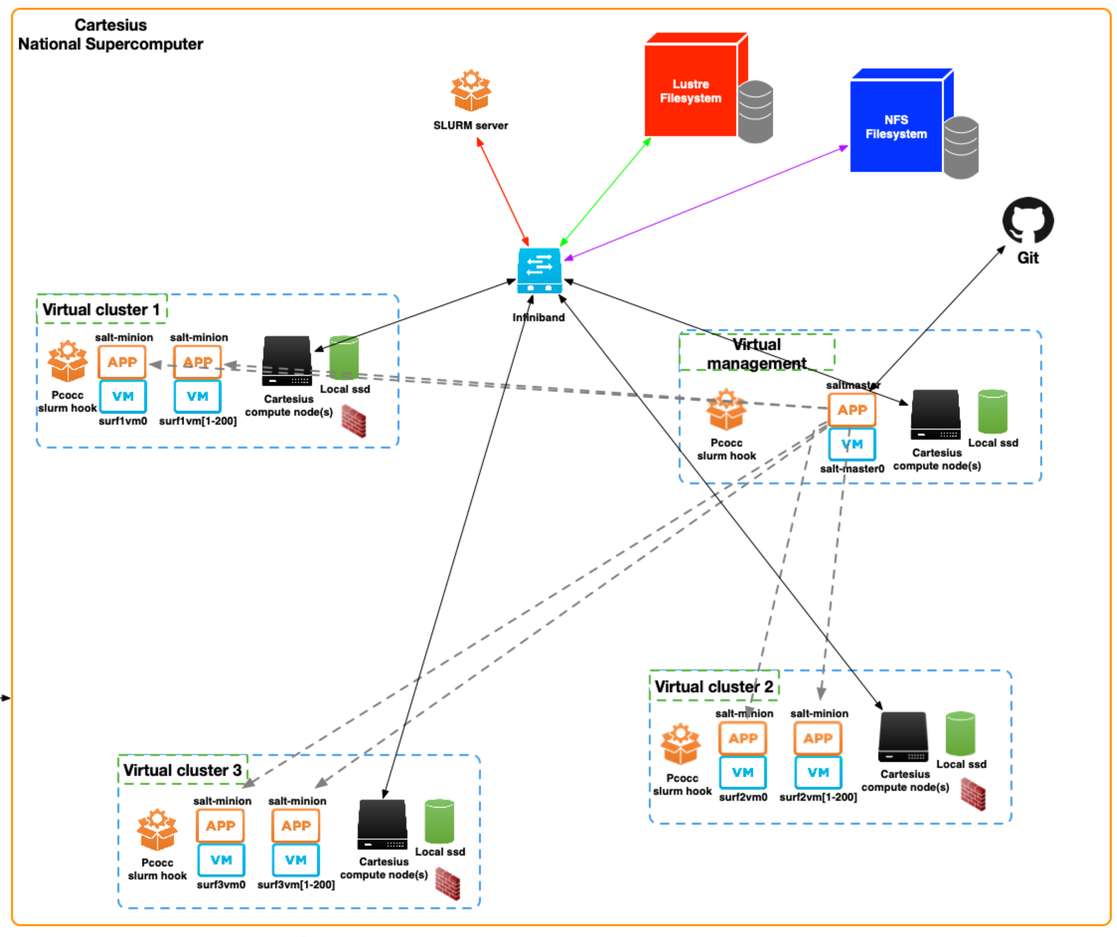
Platform as a service

SURFsara is developing a "Platform as a service" for processing sensitive data by offering virtualised private clusters. The virtual clusters are automatically provisioned on the available HPC system using PCOCC (Private Cloud on a Compute Cluster), developed by CEA (https://github.com/cea-hpc/pcocc). We have tested this setup extensively on a HPC system that consists of 2000 compute nodes connected with infiniband for high-speed, low-latency communication and storage access. The platform offers tailor-made security to meet the user's requirements, while leveraging the power of a supercomputer.



Secure access and data transfer

Users can access and transfer data to their private cluster by means of a VPN, seamlessly and securely integrating the cluster into their own private network. Stringent automated security controls make sure this VPN is the only path for sensitive data to leave the cluster.



Secure sandbox for processing data

PCOCC is a middleware that provisions virtual clusters using Linux host virtualisation technology, OpenvSwitch for network virtualisation, and integrates with SLURM for deployment. Using PCOCC and its own overlay network, secured with a firewall a sandbox is provisioned from which no data can leave. After the data processing tasks are finished the sandbox is destroyed and the data is cleared.

Future work

We are looking into secure storages options that can be connected to the platform for long term storage of sensitive data. We are also planning to perform a security audit test on the final design of the platform.

Authors: Michel Scheerman[1], Lykle Voort[1], Narges Zarrabi[1], Diederik Vandevenne[1], Sharif Islam[1]

Affiliations: [1] SURFsara (Science Park 140, Amsterdam, The Netherlands)