HNSciCloud User Groups

Photon/Neutron Science

CrystFEL
Serial Femtosecond Crystallography

The Crystfel framework is used for the technique of Serial Femtosecond Crystallography (SFX) and comprises programs for data processing, simulation and visualization. It is a part of a complex, non-redistributable software stack, which is free to use by academia and non-profit organizations.

The crystfel framework is increasingly used at various synchrotrons and FELs to analyze date from serial (femto-second) x-ray crystallography. The nature of these experiments make a cloud-based distributed pipeline particularly appealing, since the framework can fully exploit large computational resources with tunable demands. The framework is well documented and vast amount of data are readily and openly available.

Goal

A growing number of scientists performing photon science experiments have no direct storage or processing resources available to analyse data in a reasonable time. The working model within the scientific collaboration (groups of n*10) requires more 'data sharing' compared to the previous generation of experiments.

The objective of the CrystFEL use case is to run 'medium' data intensive data analysis tasks (of one of the demanding photon science experiments).

Typical initial datasets are in the 100'th of TB (with a typical image size of 6-8 MB). The processing of a single image takes ~20sec (single core). Data is formatted and accessed through HDF5.

Preconditions

» Transparent data access – local storage/ cloud storage as cache
» Container as 'task/job' format – docker
» Scheduler interface to 'local' batch system – building the 'hybrid' services/cloud
» HPC middleware i.e. MPI welcomed

HNSciCloud & EOSCPilot

This use case is derived from the EOSCpilot demonstrator having very similar CPU and Storage requirements as the initial MX (Macromolecular Crystallography) use case. The demonstrator includes in addition the ability to deploy this use case in a containerized version.

www.hnscicloud.eu | @HelixNebulaSC

European Commission

The Challenge

- » n*100k images ; up to n*100TB of data volumes
- » Complex workflow
- » Partially embarrassingly parallel
- » Partially mpi - parallel
  - » Partially serial
- » Requires (collaborative) visual inspection on the way
- » Complex Non - redistributable software
  - » Free for academia & non - profit
  - » Individual user needs to accept terms & conditions
  - » Plain CVFMS unfortunately not an option
  - » Plain Docker or Singularity hubs not an option

Benefits and impact

Photon Science – Serial Femtosecond Crystallography (SFX) is a very prominent science case in structural biology, used at Free Electron Lasers & Synchrotrons with Open Data hosted by CXIDB. Including well documented tutorials and is computationally challenging

Procurer sponsoring the use case: DESY

The "Stiftung Deutsches Elektronen-Synchrotron DESY" is one of the world's leading laboratories for particle physics and photon science. As a member of the Helmholtz Association in Germany, DESY is a non-profit research organization funded by public funds from the government and the states.

DESY has a long-standing experience in the design and operation of accelerators for high energy physics and synchrotron radiation research. DESY today operates the most brilliant synchrotron radiation source worldwide (PETRA III), two Free