



Scientific data analysis executed in true hybrid clouds - Serial Femtosecond Crystallography/CrystFEL

Martin Gasthuber, Sergey Yakubov

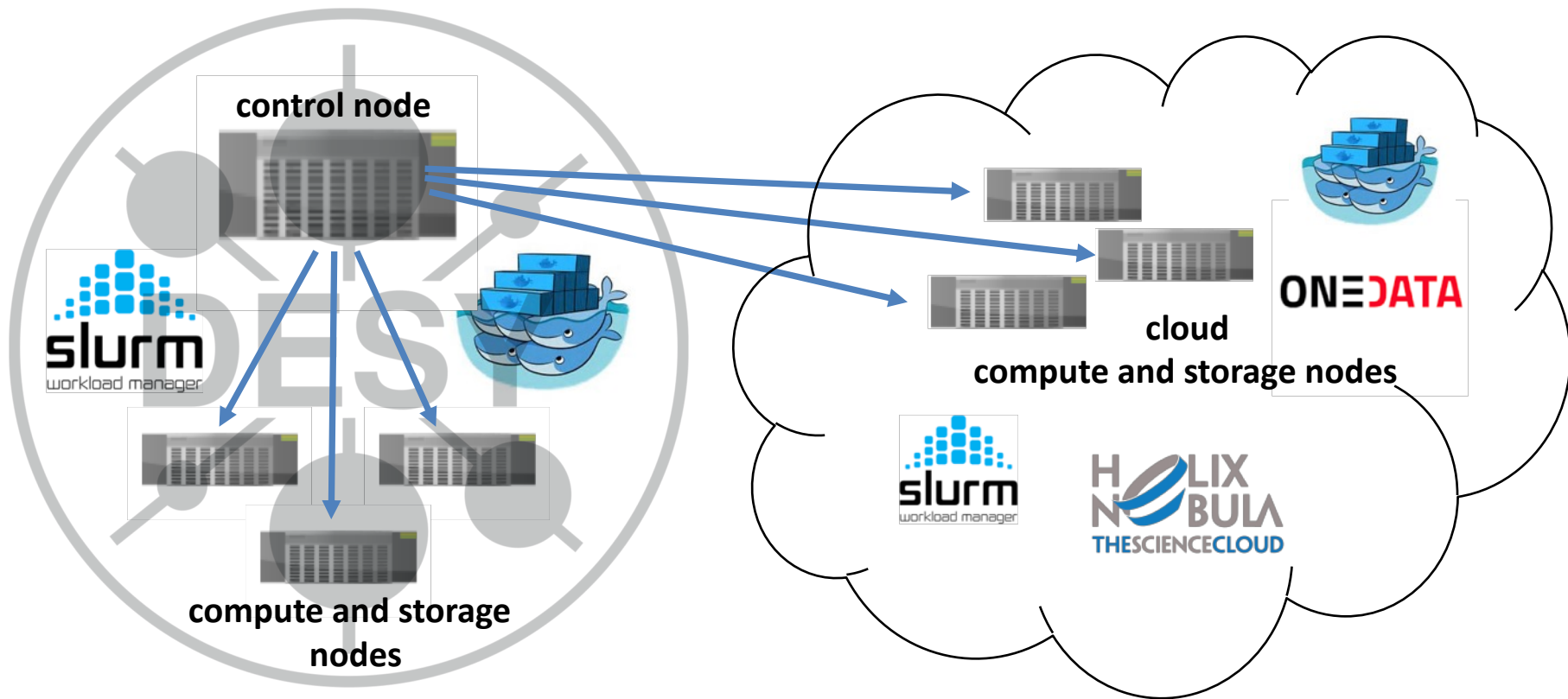
DESY IT



Hybrid Cloud at DESY

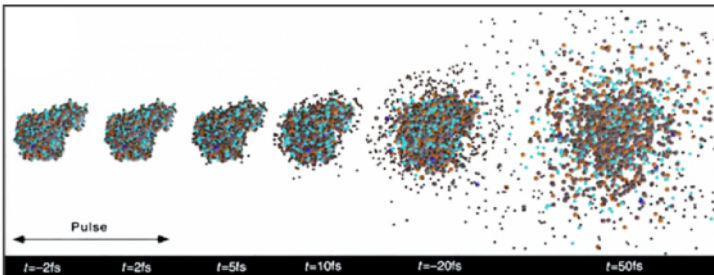
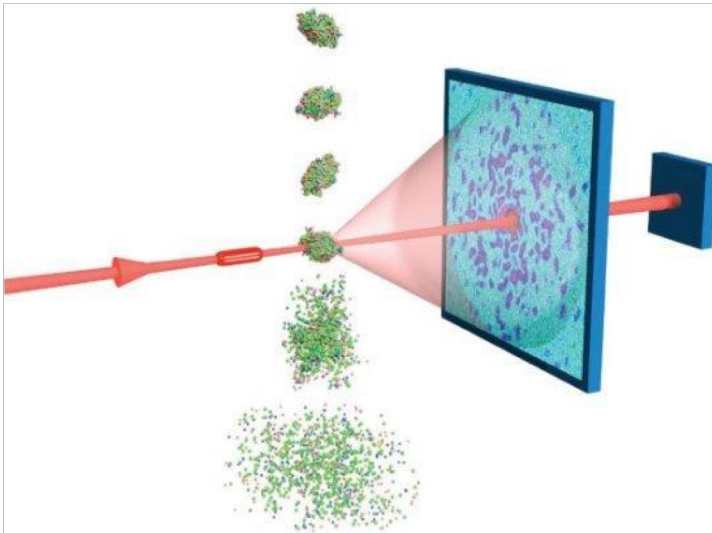


- Resource manager (SLURM) used at DESY is extended to submit jobs in the cloud
 - Cloud VMs are automatically allocated and deallocated
 - OneData provides transparent data access between cloud and local storage
 - User does not need to know where his job is executed / adjust it to cloud infrastructure





Use case – Serial Crystallography/ CrystFEL



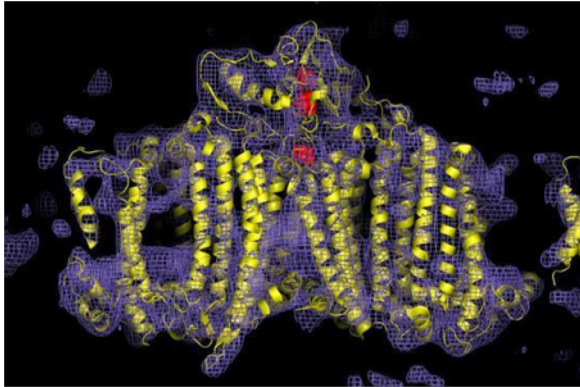
- ☞ Photon neutron science demonstrator for EOSC Pilot
- ☞ Molecular 3D imaging at atomic scale
- ☞ Very high time resolution
- ☞ Resolve inner structures (not just surface) of biological objects
- ☞ To reconstruct a 3D image a large amount of random oriented single images have to be analysed

source: https://cid.cfel.de/research/femtosecond_crystallography/

<http://www.desy.de/~twhite/crystfel/index.html>



Example: Protein Complex Photosystem



Reconstructed image

- Biological factory in plant cells
- converts sunlight to energy during photosynthesis
- 3,000,000 images have been collected during experiment
- after initial reduction 10,000 images need to be analyzed
- 6.6 MB image size, 20 sec/image compute time
- 20 TB data (before reduction) / 66 GB useful data/ 2 days simulations on a single CPU

Executing analysis locally or in hybrid cloud

