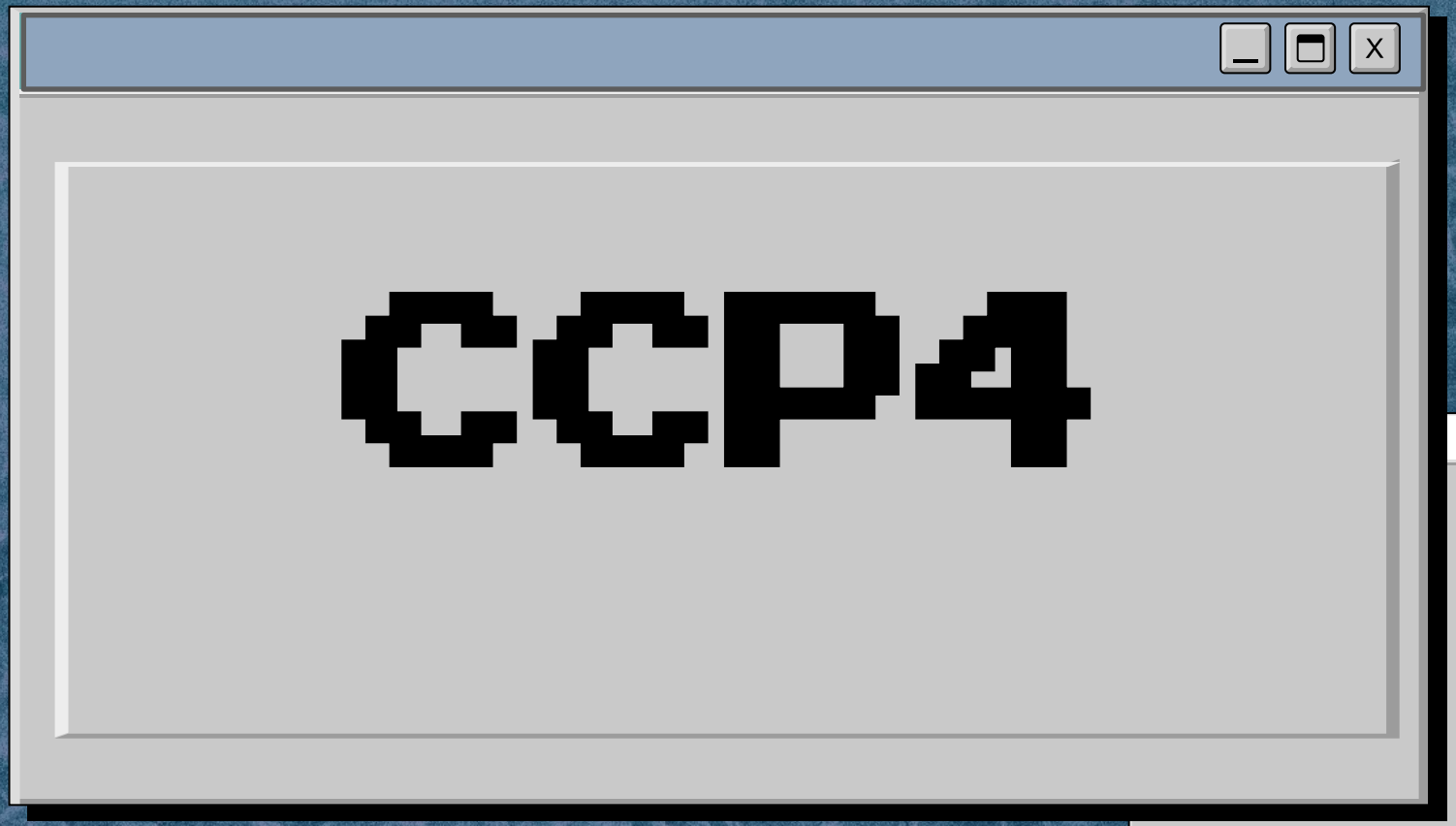
Maria Fando

Delivering HPC Power for Structural Biologists with CCP4 Cloud

Computing Insight UK 2023

07/12/2023


CCP4

"Collaborative Computational Project No. 4 in Protein Crystallography"

CCP4

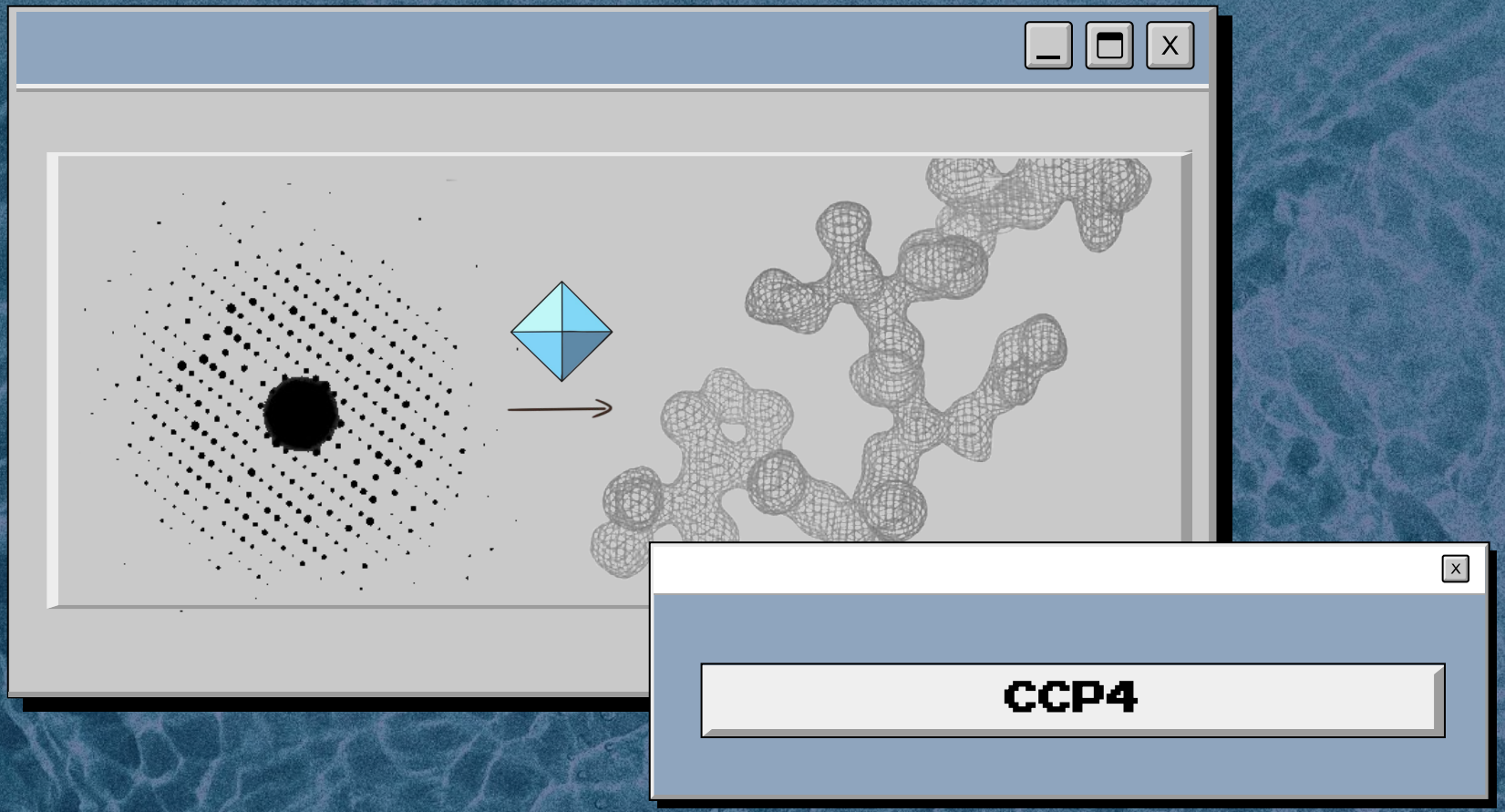
CCP4 was set up in 1979 to support collaboration between researchers working on MX software in the UK, and to assemble a software suite to satisfy the computational requirements of the relevant UK groups

CCP4

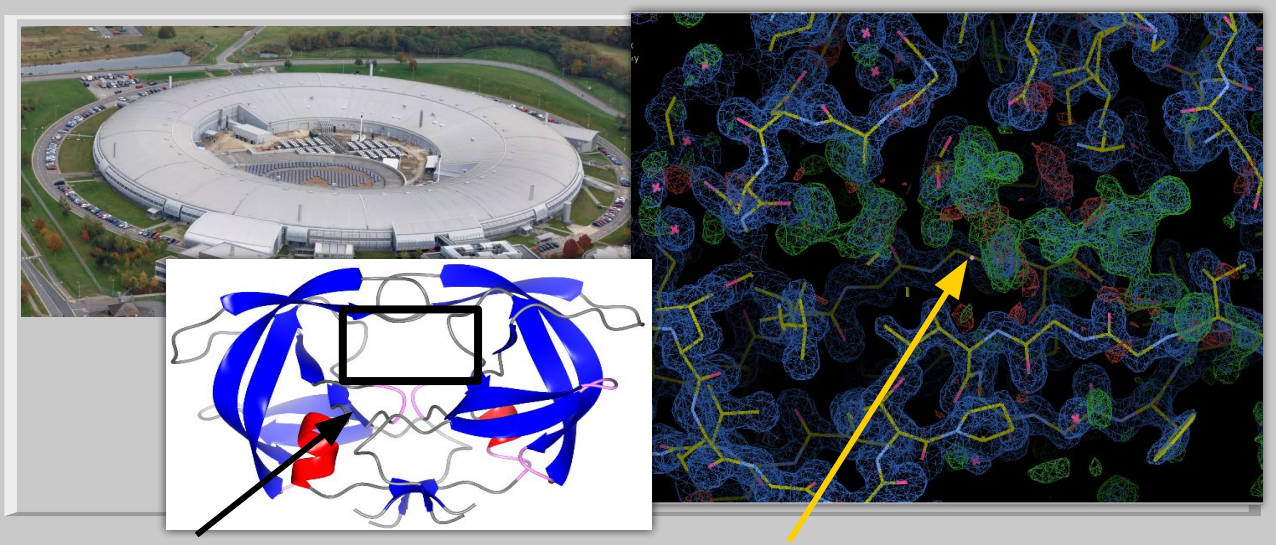
The CCP4 Suite is a comprehensive suite of software for macromolecular crystallography. It contains about 700 program components

CCP4

CCP4 user's map


X
X
CCP4

diamond



Active Site

Drug inhibitor in crystal

Data in Crystallography

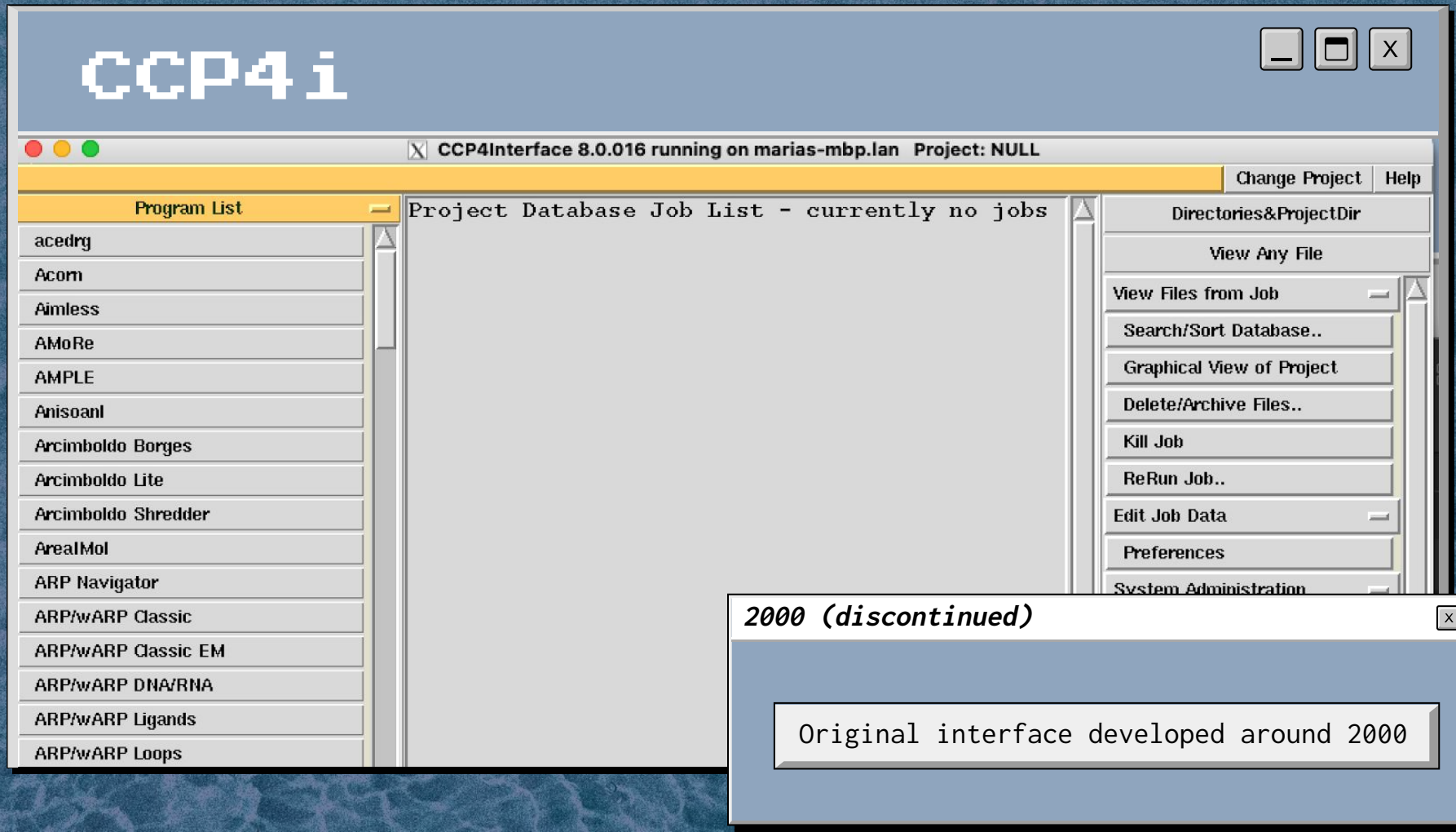
Progress in the determination of three-dimensional macromolecular structures from diffraction images is achieved partly at the cost of increased computational time and data volumes

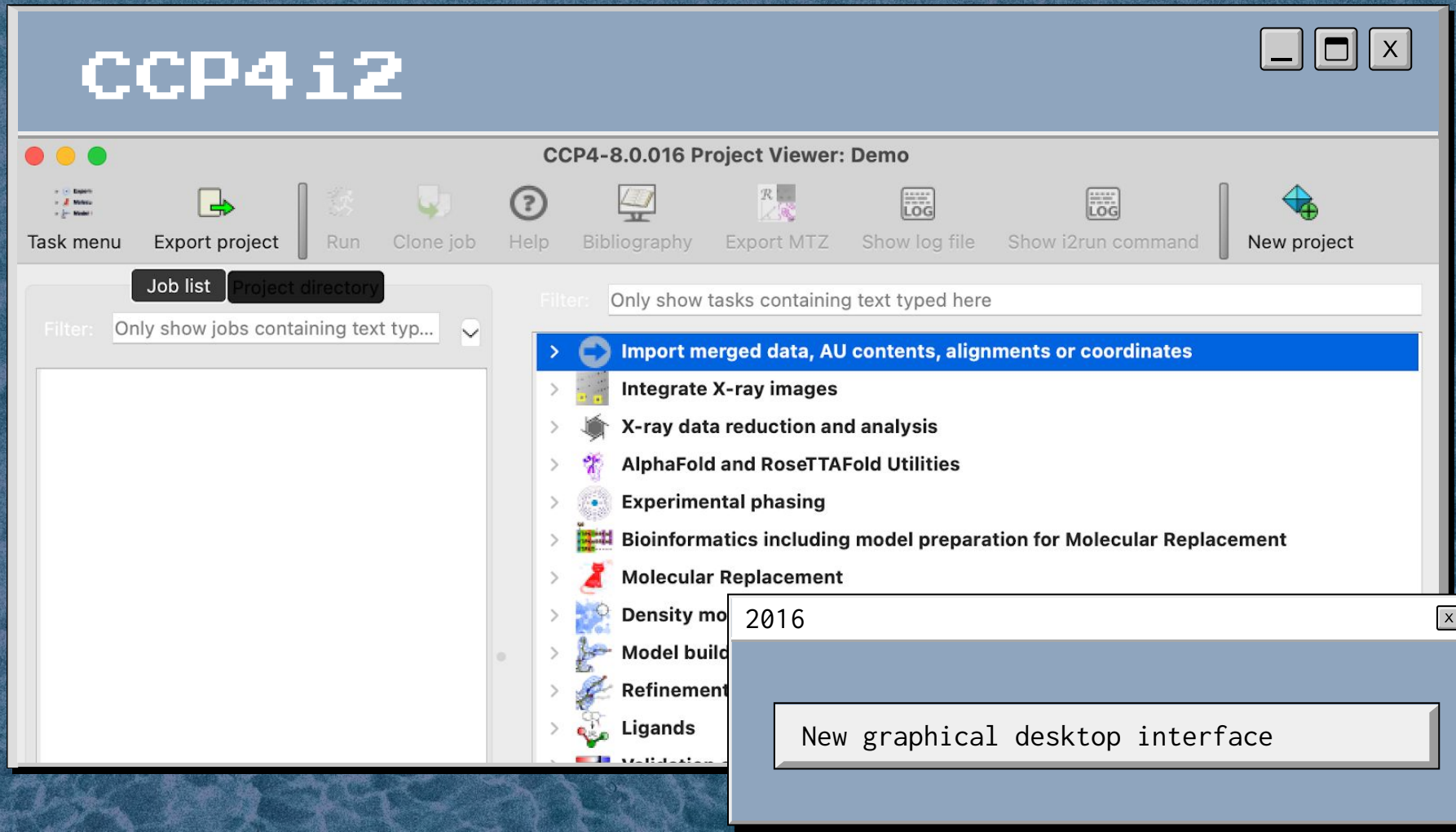
CCP4 shifted heavily to an automatic structure solution at the cost of higher CPU demand

Combination of data, AI and computing power starts giving a boost for automation

Recent example: Structure determination using predicted models

- unprecedented accuracy in predicting protein structures in 3d
- expanding proteomics to genomics scales (from 180K known to 200M structures in AlphaFold (by DeepMind) and 772M in ESM (by Meta) datatbases)


CCP4i
CCP4Interface 8.0.016 running on marias-mbp.lan Project: NULL
Change Project
Help
Program List
acedrg
Acorn
Aimless
AMoRe
AMPLE
Anisoanl
Arcimboldo Borges
Arcimboldo Lite
Arcimboldo Shredder
ArealMol
ARP Navigator
ARP/wARP Classic
ARP/wARP Classic EM
ARP/wARP DNA/RNA
ARP/wARP Ligands
ARP/wARP Loops
Project Database Job List - currently no jobs
Directories&ProjectDir
View Any File
View Files from Job
Search/Sort Database..
Graphical View of Project
Delete/Archive Files..
Kill Job
ReRun Job..
Edit Job Data
Preferences
System Administration
2000 (discontinued)
Original interface developed around 2000


CCP4i2
CCP4-8.0.016 Project Viewer: Demo
Task menu
Export project
Run
Clone job
Help
Bibliography
Export MTZ
Show log file
Show i2run command
New project
Job list
Project directory
Filter:
Only show jobs containing text typ...
Filter:
Only show tasks containing text typed here
Import merged data, AU contents, alignments or coordinates
Integrate X-ray images
X-ray data reduction and analysis
AlphaFold and RoseTTAFold Utilities
Experimental phasing
Bioinformatics including model preparation for Molecular Replacement
Molecular Replacement
Density mo
Model build
Refinement
Ligands
2016
New graphical desktop interface

CCP4 Cloud

cloud.ccp4.ac.uk

My Projects/Demo

54% 0%:1% maria

Open | Add | Rename | Clone | Move | Delete | Export | Import | Join | Tutorials | Help

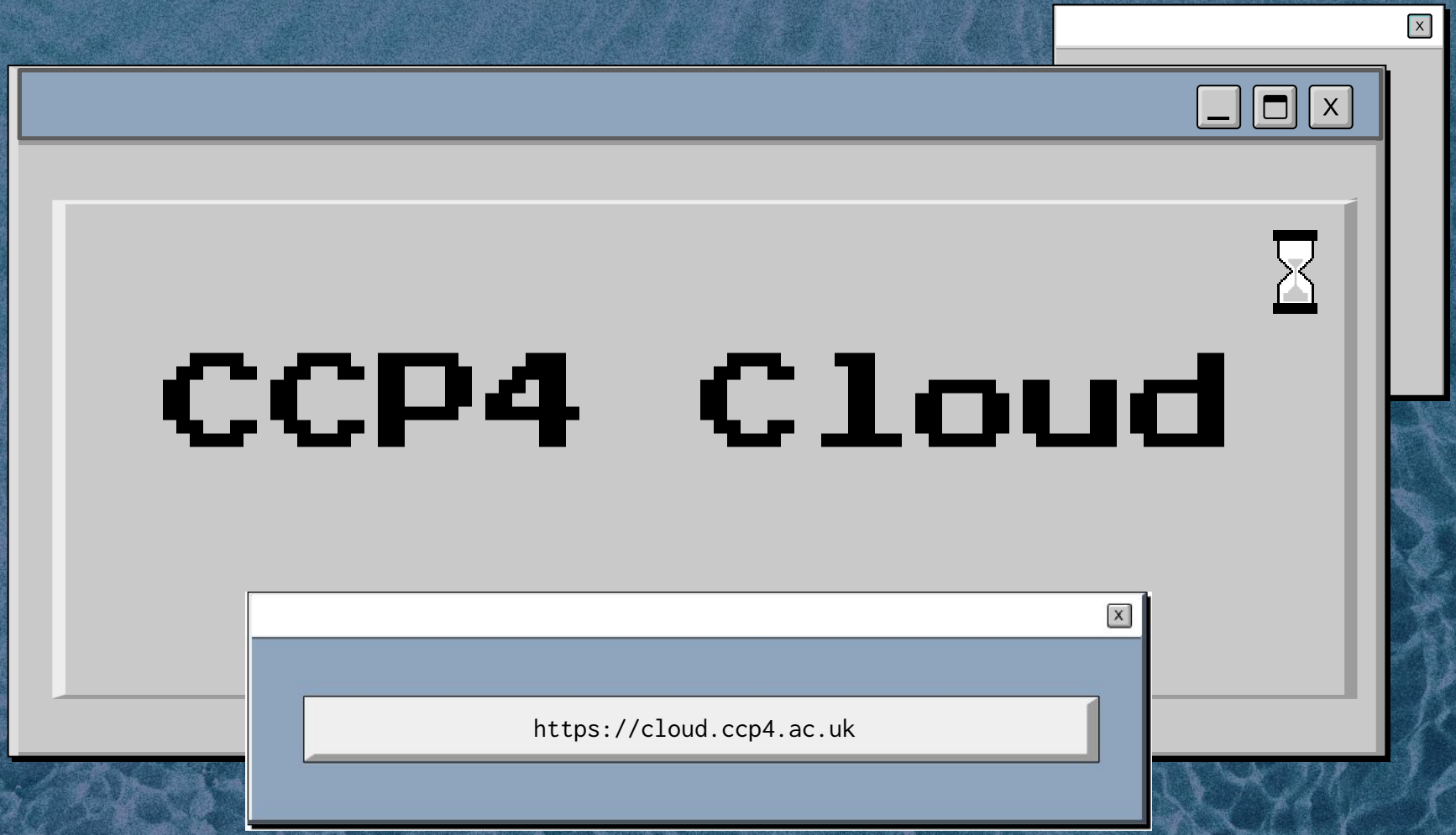
ID	Name	R_{free}	Disk (MBytes)	CPU (hours)	Date Created	Last Opened

There are no projects in folder "My Projects/Demo".

Use "Add" button to create a new Project;
"Import" button for importing a project exported from CCP4 Cloud;
"Join" button for joining project
or "Tutorials" button for loa
or click on page title or folder

2018

browser based


CCP4 Cloud
https://cloud.ccp4.ac.uk

CCP4 Cloud Initiative

Conceived in 2016

- Funded by BBSRC UK and CCP4

Response to demands and trends rapidly emerging in the field

- CPU power (due to increased automation)
- Centralised database support (due to expansion of methods based on data templates)
- Software as a service (due to increased size and complexity of software setups)
- Supporting distributed projects for team work
- Cloud model for geographically-agnostic access and project data safety
- Supporting personal mobile platforms (tablets and smartphones)
- Communication with data facilities (synchrotrons, PDB, AFDB, etc)

Availability

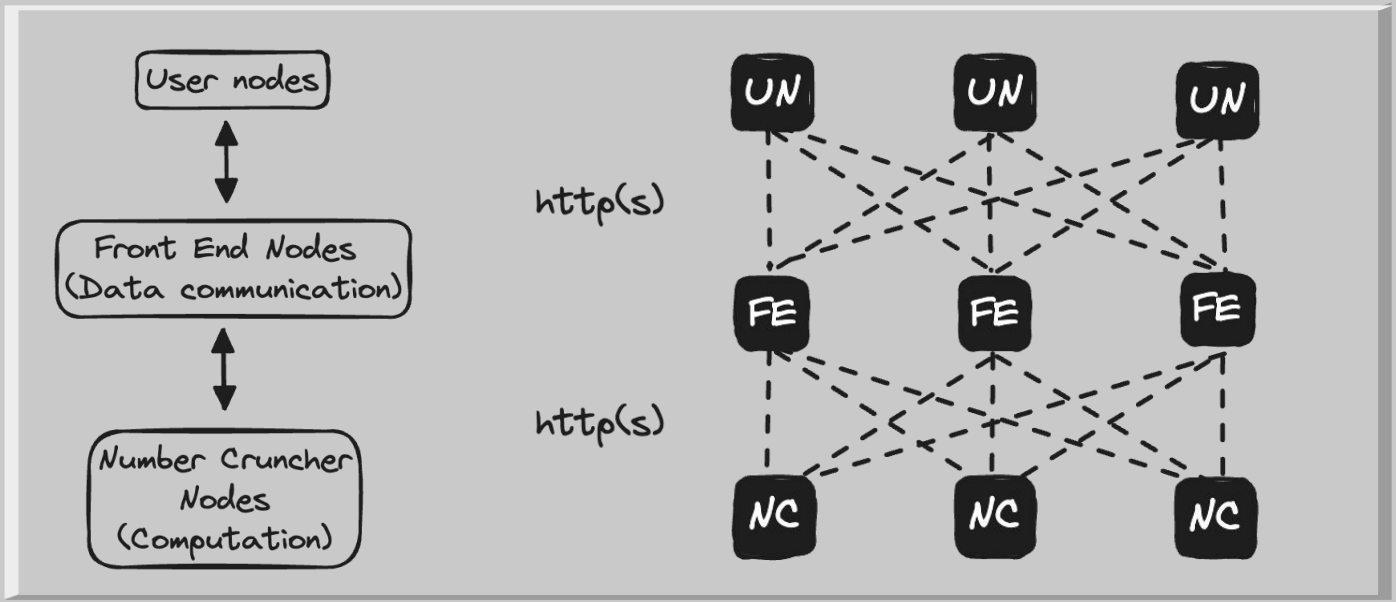
https://cloud.ccp4.ac.uk

https://www.ccp4.ac.uk/download

Part of standard CCP4 distribution package starting from 7.1 release series (since June 2020)

CCP4 Cloud Architecture

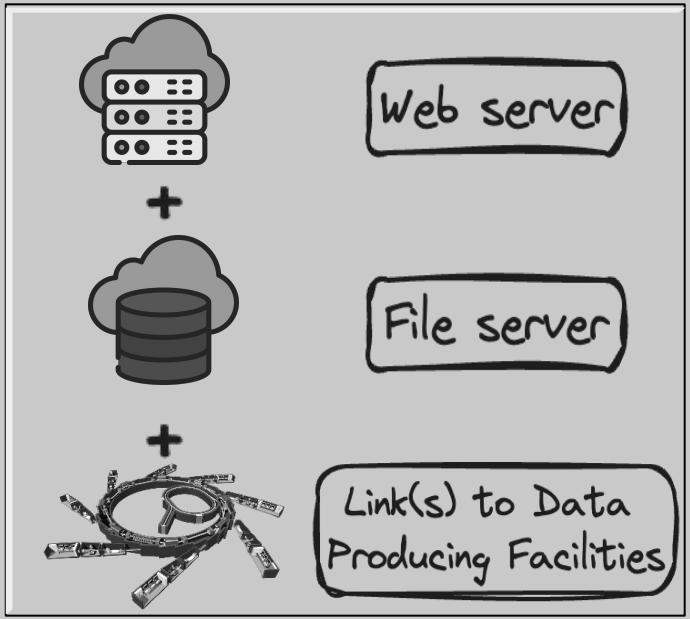
CCP4 Cloud Architecture



CCP4 Cloud Architecture

Front End Nodes:

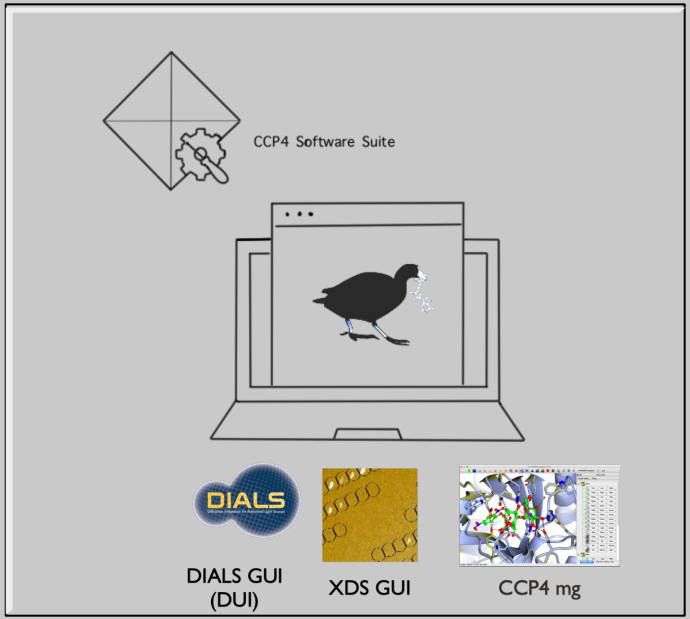
- provide all data logistics in the Cloud
- represent web-servers with storage for user data and projects
- may obtain data from Data Producing Facilities (experimental at the moment)
- do not run any calculations
- despatch jobs to Number Cruncher Nodes



CCP4 Cloud Architecture

CCP4 Cloud Client:

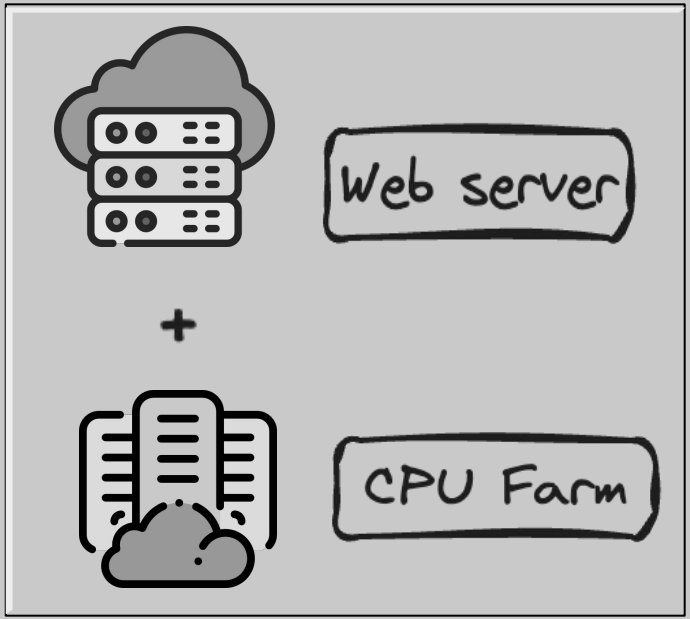
- local server which effectively makes user's device a part of CCP4 Cloud, proprietary to that user
- used to run interactive desktop applications, and also image processing where image data cannot be placed in the Cloud
- installs out-of-box as a CCP4 package



CCP4 Cloud Architecture

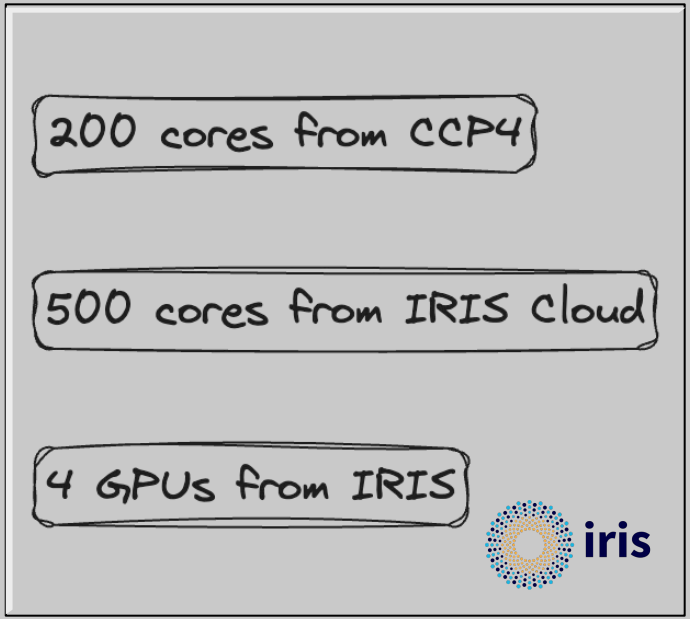
Number Cruncher Nodes:

- only run calculations
- receive jobs from Front End Nodes and send results back to the sender FE
- may be placed on a single or multiple hardware hosts

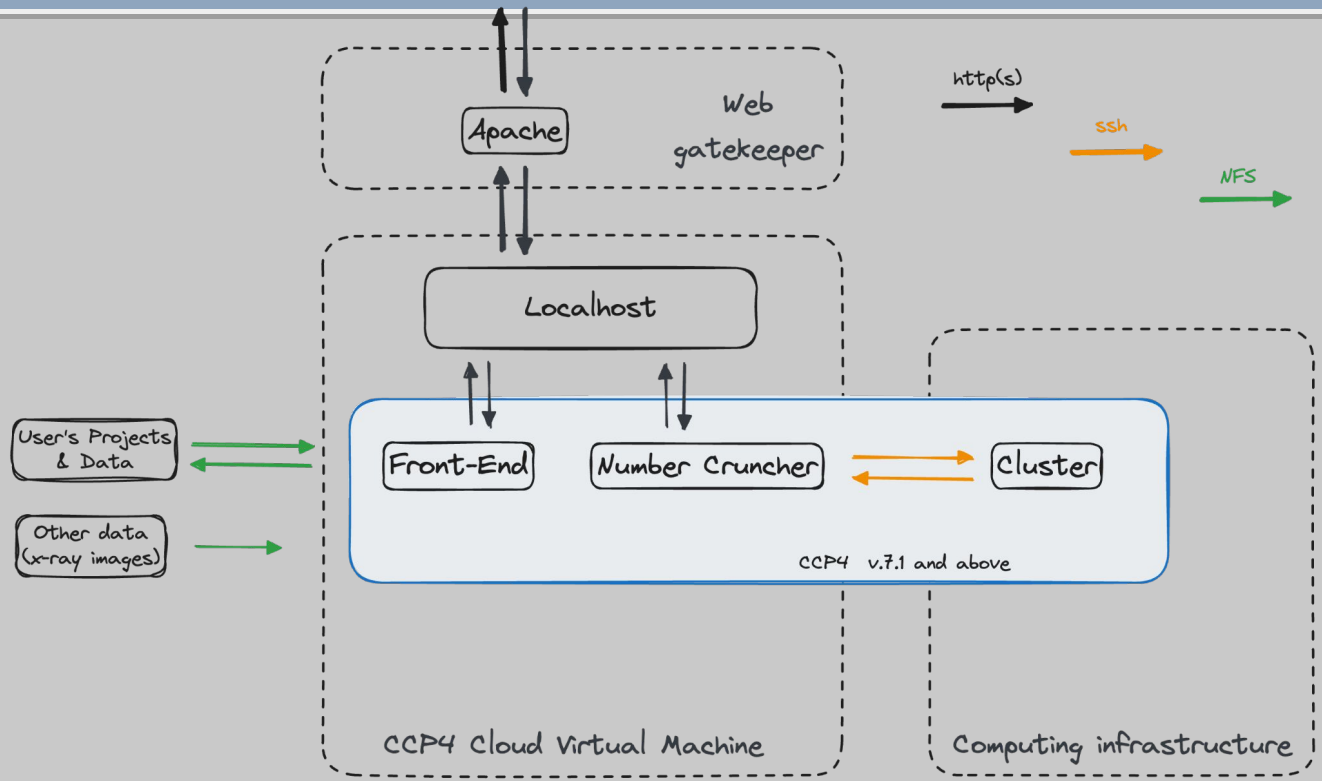


CCP4 Cloud Architecture

- Computing back-end can be a cluster (SLURM, GRID engine, etc) or queue-less system
- NC Web-Server and back-end can be on same or different machine



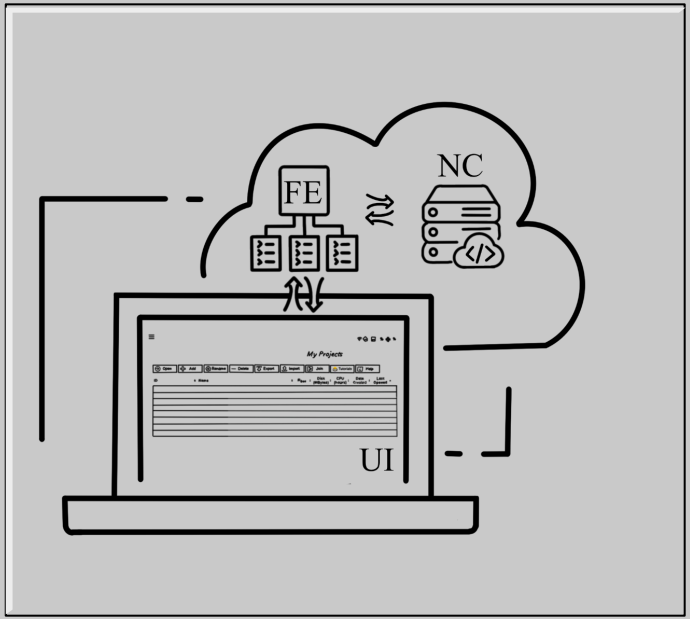
Communication protocols



CCP4 Cloud configurations

CCP4 Cloud configurations

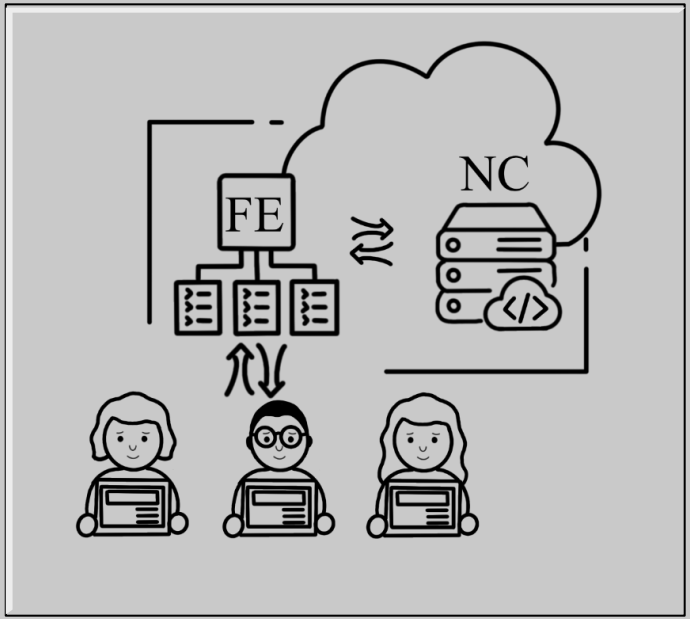
Single-host configuration, suitable for an individual working without need for internet connection



FE - Front End; NC - Number Cruncher; UI - User Interface

CCP4 Cloud configurations

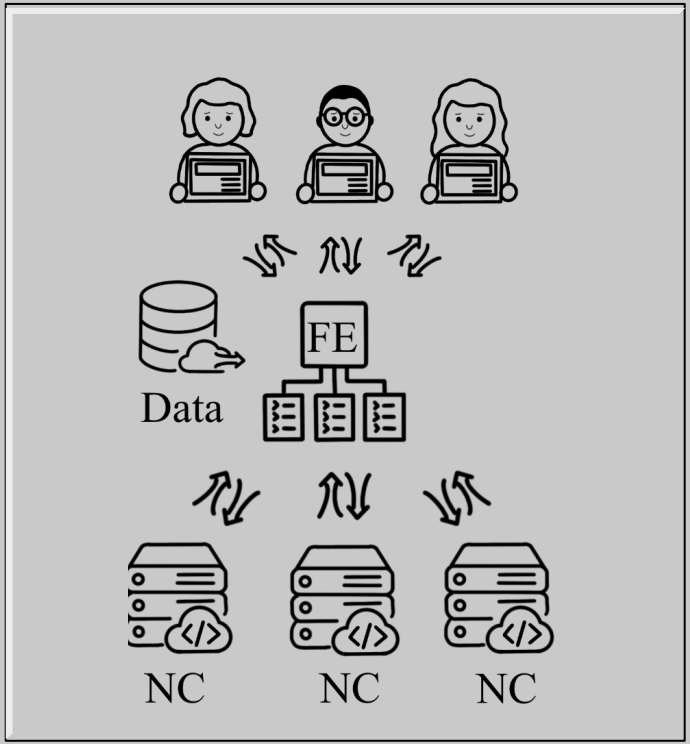
A multi-user setup using a central host machine, suitable for small to medium-sized laboratories.



FE - Front End; NC - Number Cruncher

CCP4 Cloud configurations

Fully distributed, multi-component setup with single point of access, suitable for large facilities and research centres



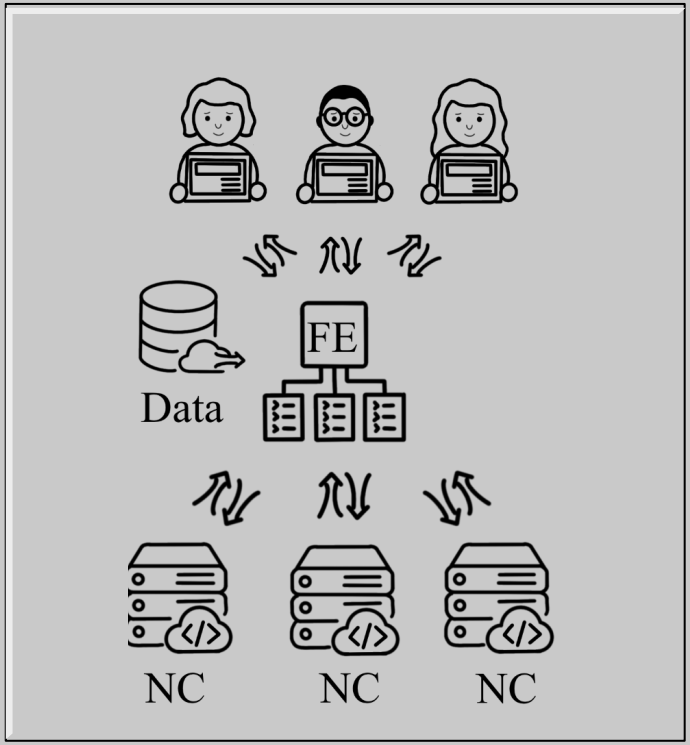
FE - Front End; NC - Number Cruncher

CCP4 Cloud configurations

Fully distributed,
multi-component setup with
single point of access,
suitable for large facilities
and research centres

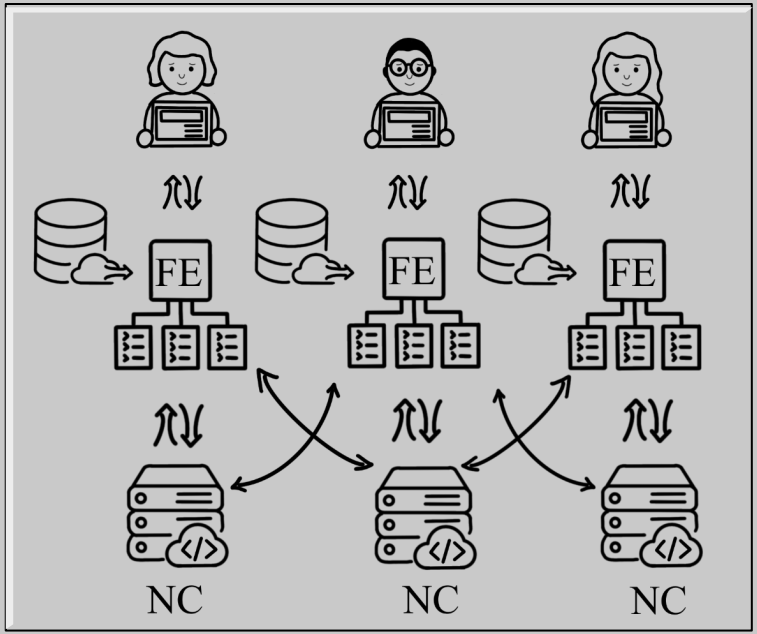
https://cloud.ccp4.ac.uk

FE - Front End; NC - Number Cruncher



CCP4 Cloud configurations

Fully distributed, multi-component setup with multiple points of access, suitable for large facilities and research centres



FE - Front End; NC - Number Cruncher

Implementation details

- All server nodes are based on the Node JS platform
- Browser side: HTML5, WebGL, custom Javascript widget framework based on jQuery, jQuery-UI and React
- Job launching framework: Python
- Job report framework: RVAPI (dynamic web content) from CCP4
- Job workflow framework based on abstract task and data models
- Update mechanism
- Script-assisted installation, auto-setup for CCP4 Cloud Client
- No principal restrictions on the number and location of computational nodes

CCP4 Clouds instances

Main CCP4 Cloud instance at CCP4-Harwell from 2018:

- - Over 4,500 user accounts
- - Over 100,000 jobs/year

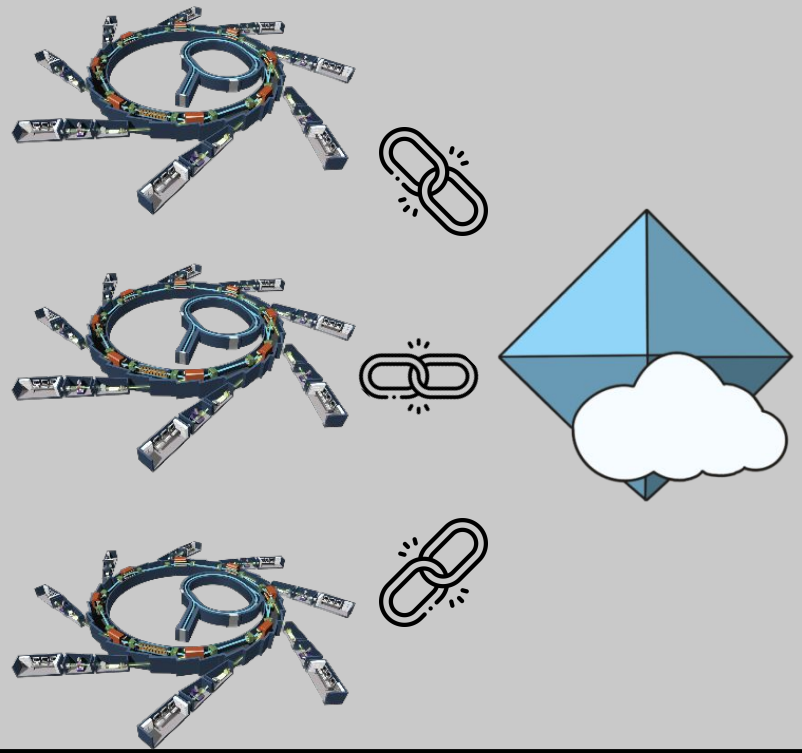
CCP4 Cloud instances at partner sites, including industrial sector:

- EMBL (Hamburg)
- Francis Crick Institute (London)
- Newcastle University
- University of Exeter
- Incyte Inc (Virginia, USA)

Future plans

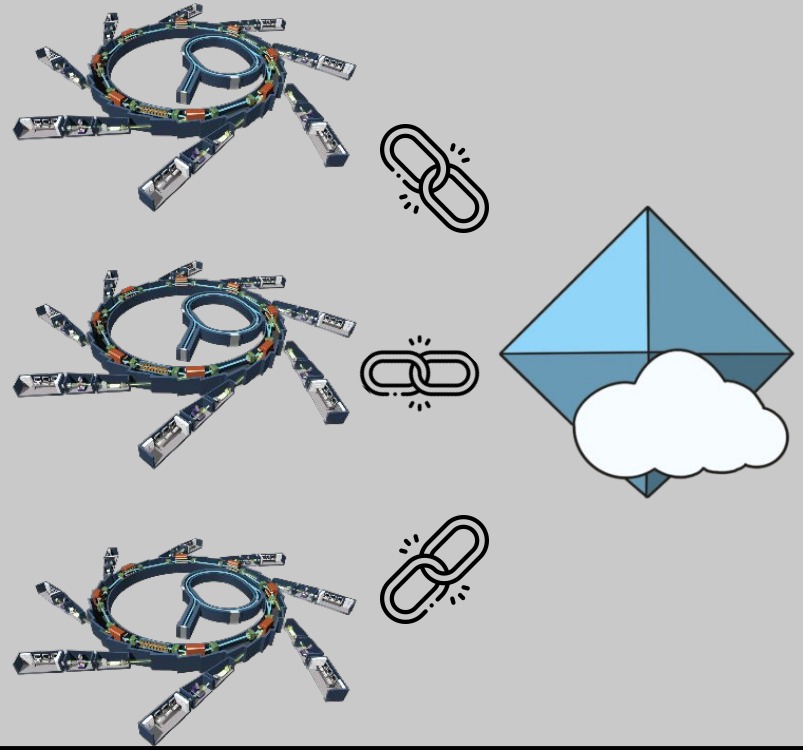
Future plans

Make data links between experimental facilities and in-house X-ray diffractometers and CCP4 Cloud



Future plans

Make data links between experimental facilities and in-house X-ray diffractometers and CCP4 Cloud

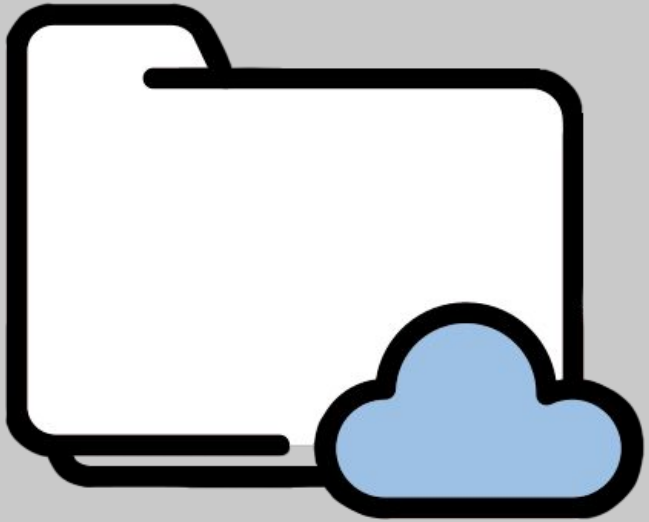


ada lovelace centre

Future plans

CCP4 Cloud Archive:

- Develop
- Maintain
- Popularise



Conclusions

CCP4 Cloud

Mitigates software complexity

- ➔ Supporting wide variety of computing platforms is difficult
- ➔ Full installation with 3rd party databases and software is difficult

Meets methods and software demands

- ➔ Modern automatic methods require more CPU and memory than most local setups can afford

Facilitates data logistics and distributed team working

- ➔ Growing volumes of data from modern sources are difficult to handle locally
- ➔ File exchange in distributed collaborations is usually a mess

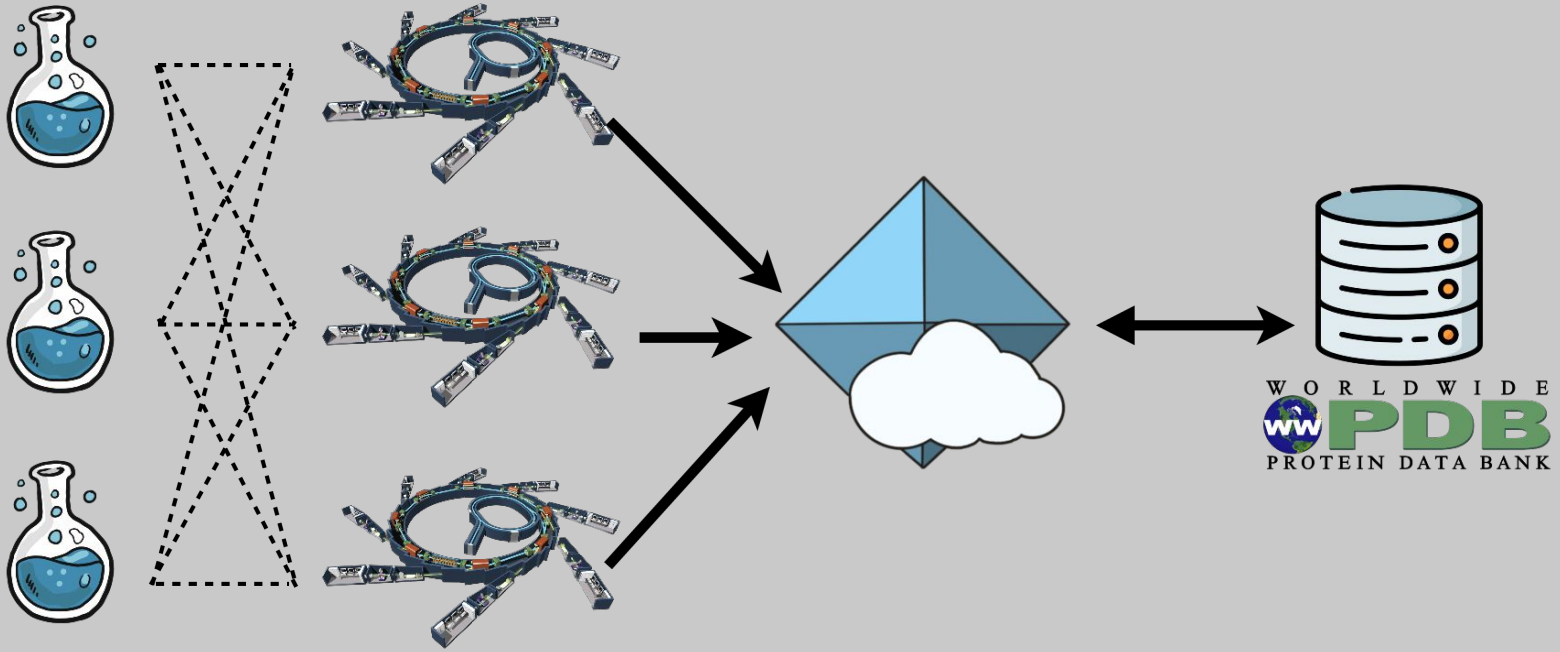
Provides for data security and retention

- ➔ Increasingly more difficult to distribute for modern systems and corporate environments
- ➔ Cloud solutions are safer, getting preferential in industry
- ➔ The lifetime of data stored in the cloud is considerably longer (effectively infinite) than that usually achieved with locally maintained hardware

CCP4 Cloud's key features

- Software, resources and data as a service: go-and-use
- Cross-platform compatibility: can be used on Windows, Linux, Mac OSX, tablets and smartphones
- Rich project development functionality
- All stages of structures solution: from image processing to PDB deposition
- Integrates access to web-resources such as PDB and AFDB
- Facilitates teamwork by sharing projects in real time with various levels of access
- Can be run locally
- Can be installed in a lab, institute or firm
- Highly configurable and adaptable to using mixed distributed computational resources
- Integrated documentation and tutorials

Data Production in Structural Biology



Research Labs → Imaging Facilities → Structure Solution → Data Bank → Utilisation

Acknowledgments

STFC, CCP4, Harwell, UK:

Eugene Krissinel, Andrey Lebedev, Oleg Kovalevskyi, Ronan Keegan, Charles Ballard, Ville Uski, Jools Wills

Newcastle Uni UK:

Martin Noble

Uni Southampton UK:

Ivo Tews

MRC/LMB, Cambridge, UK:

Robert Nicholls

EMBL-EBI, Hinxton, UK:

John Berrisford

Uni Leiden, The Netherlands:

Navraj Pannu, Pavol Skubak

Global Phasing Ltd, Cambridge, UK:

Marcin Wojdyr, Clemens Vonrhein

Uni York, UK:

Stuart McNicholas, Filomeno Sanchez Rodriguez, Paul Bond

Uni Liverpool, UK:

Adam Simpkin, Jens Thomas

Uni Birmingham, UK:

Christopher Oliver

CCP4, STFC & RCaH

CCP4 Collaboration, CCP4 developers

CCP4 Cloud users
Worldwide

CCP4 School hosts

Ed Lowe
Oxford University

Andy Purkiss
Francis Crick Institute, London

Grzegorz Chojnowski
EMBL-Hamburg

Arnaud Basle
Newcastle University

Michael Isupov
University of Exeter

Biotechnology and Biological Sciences Research Council *(BBSRC) UK*

UKRI Research Complex at Harwell

iris

CCP4

UKRI Science and Technology Facilities Council

Newcastle University

THE FRANCIS CRICK INSTITUTE

EMBL European Molecular Biology Laboratory

UNIVERSITY OF EXETER

ada lovelace centre

UKRI Biotechnology and Biological Sciences Research Council

Research grant BB/L0070317/1 (2014-2019)