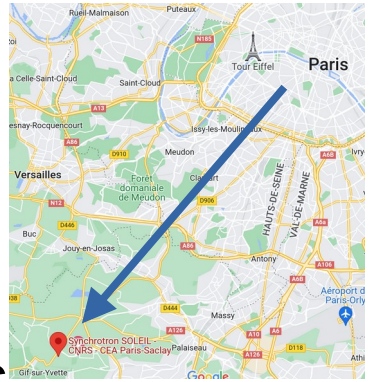
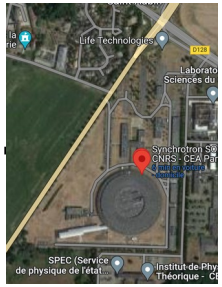
DARTS on-demand computing : heavy-duty data treatment for all – Emmanuel Farhi

Our group

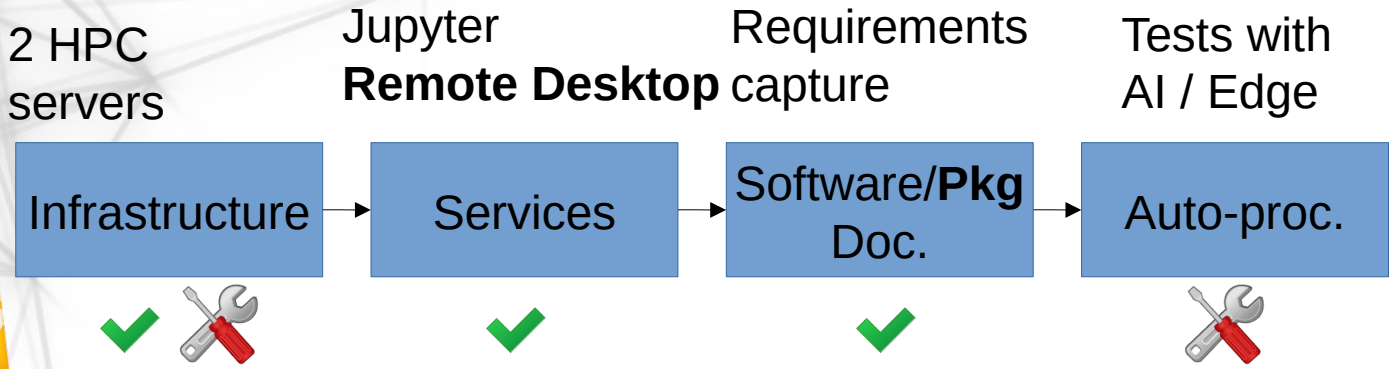
- Synchrotron SOLEIL, close to Paris.
- *Groupe de Réduction et d'Analyse de Données Expérimentales de SOLEIL* (**GRADES**).
- Created in Feb 2020, in the EXP Div.
- Staff: 6 perm+4 temp.
- Topics: scientific software (diff, spectro/abs, simulation, ...), Debian/Ubuntu packaging, deployment, coding, documentation, support, …





Our Tasks and Roadmap

- Help beam-line users to better treat the data and increase the scientific output.
- Support beam-line scientists.
- Provide access to scientific software.



DARTS

stands for “Data Analysis Remote Treatment Service”.

In short:

- A virtualization service that works on the web (displayed in your browser)
- Installed and configured in a few minutes
- Works on a distributed network of machines
- Supports GPU
- Allow sharing of opened sessions

Open-source code:

- https://gitlab.com/soleil-data-treatment/soleil-software-projects/remote-desktop
- https://packages.debian.org/sid/qemu-web-desktop

Why such a service ?

Facts:

- Users wish to access data and software, with minimal effort.
- Jupyter Notebooks do not fulfill all user profiles.
- Some apps require a proper display (GUI).
- IT services thrive to minimize the infrastructure workload (hardware and software). Better to centralize the computing resources.
- An HPC cluster does not suit common user needs for computations.

Solutions exist:

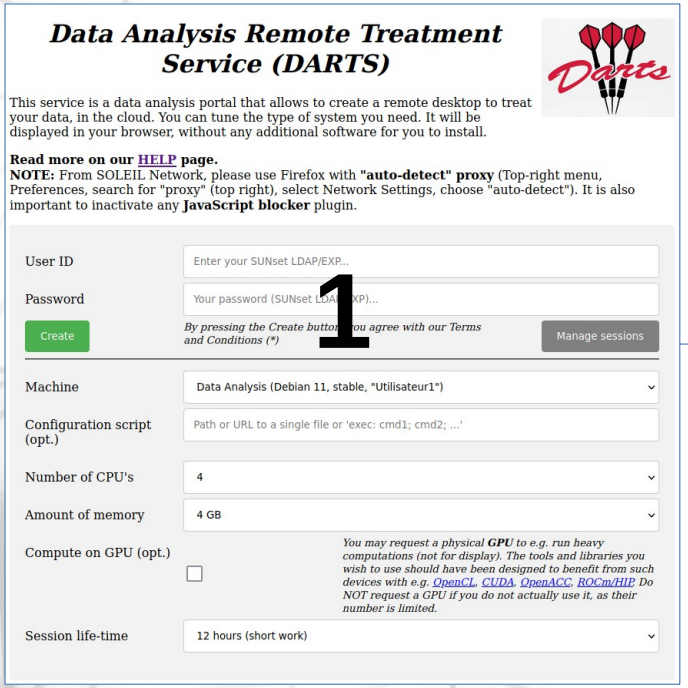
- Docker.
- Microsoft/Amazon/Google provide VM solutions.
- CernVM (over OpenStack) https://cernvm.cern.ch/.
- VISA/OpenStack, promoted by ExPaNDS/PaNOSC.

Best effort:

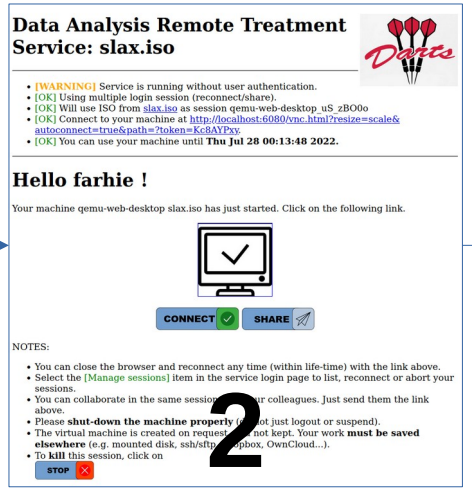
Our group is limited in size, VISA was not ready at SOLEIL when we created a simple similar solution:

- DARTS was designed in a fortnight.
- Provides similar user experience.
- fits in 1600 LOC with simple technologies.

User experience

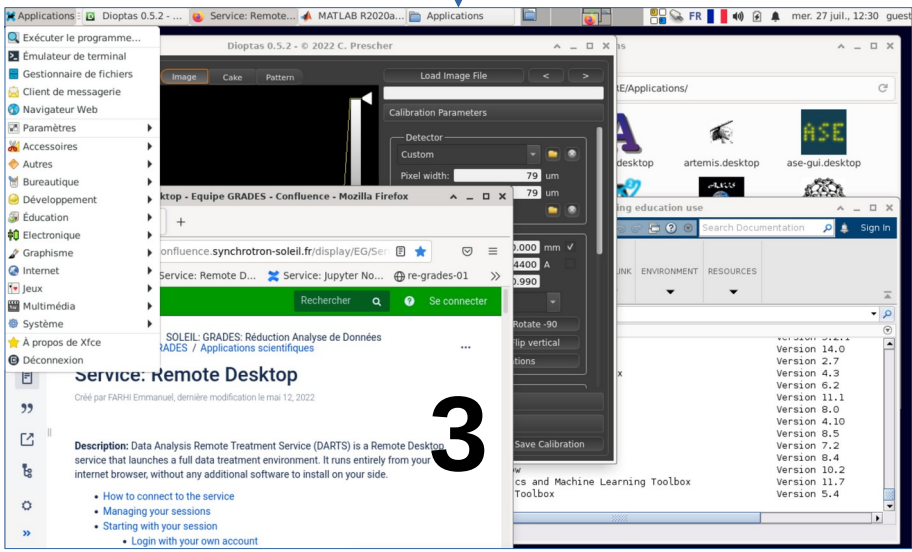


(10 s)



Connect

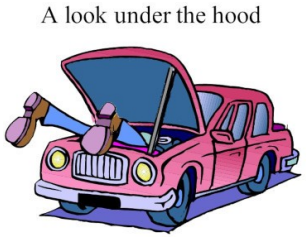
Work

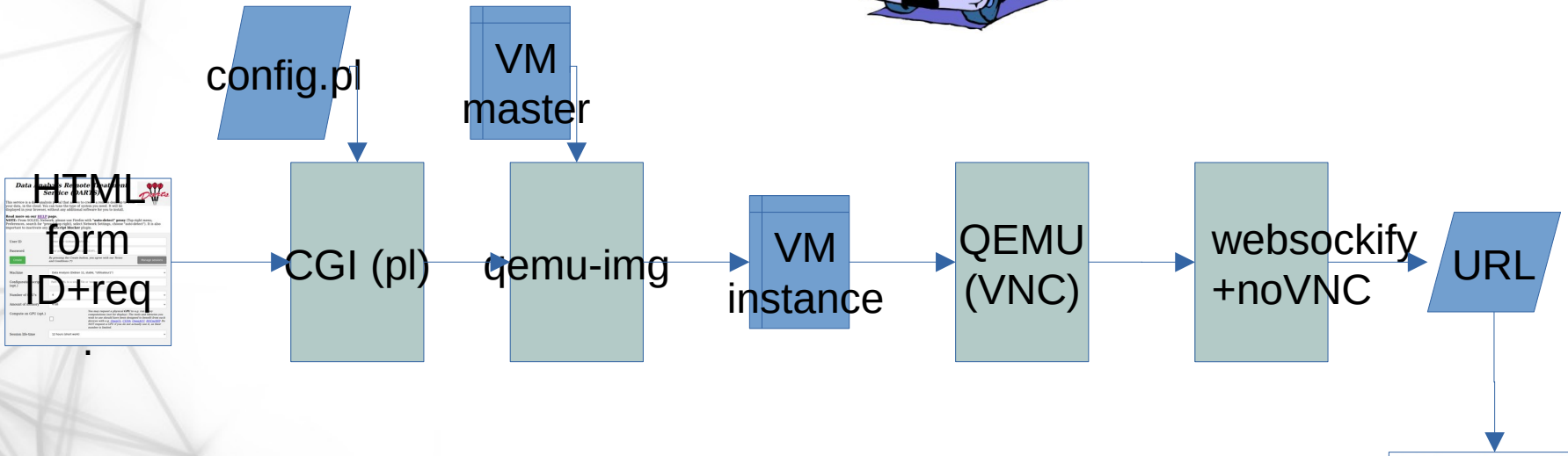


http://server/qemu-web-desktop
Authenticate, and select what resources you need.

No need to install anything locally.
Direct access to computing resources.

Under the hood





1) An HTML form collects information.
2) Calls a CGI Perl script.
3) A VM snapshot is created from a master.
4) The VM snapshot is booted with requested CPU/mem/GPU directly with QEMU.
5) QEMU attaches a VNC to the VM, and forwards it to the browser via websockify+noVNC.
6) A monitoring page allows to view/abort sessions.

Installation / configuration

Requirements:

- Apache
- Perl CGI
- QEMU
- NoVNC

Installation:

- `sudo apt install qemu-web-desktop`
- `make; sudo make install`

Configuration

`/etc/qemu-web-desktop/config.pl, /usr/share/qemu-web-desktop/html/desktop/index.html`

1) Specify identification scheme (LDAP, SMTP/IMAP, email), load thresholds, proxy, etc … (in `config.pl`).
2) Activate load-leveller among available machines, when available (in `config.pl`).
3) Set available CPU/mem/life-time choices (in `index.html`).
4) Activate GPU support if needed (in `index.html`).
5) Activate user customization at boot, if desirable (in `index.html`).

Preparing VM's for DARTS

Adding environments

`/etc/qemu-web-desktop/machines.conf; /var/lib/qemu-web-desktop/machines`

ISO, QCOW2, VMDK, VDI are supported.
You can build the VM's by hand or automatically (below).
Copy files/specify URLs in `machine.conf`, run '`qwdctl download`'

→ The HTML login form is automatically updated.

Automated VM generation (10 min)

- Build our environments for production using scripts at https://gitlab.com/soleil-data-treatment/infra-config
- Feed the environments with 100+ software, using Deb packages.
- Add more specific software (Matlab, AlphaFold, Fiji, ...)

DARTS at SOLEIL

Installed it on a set of servers (e.g. 320 cores, 2 TB mem, ~100 TB disk, 14 GPU's – 50 k€).

Prepared environments at Synchrotron SOLEIL

We contribute to Debian Science packages, and deploy these.<
https://salsa.debian.org/pan-team/soleil-packaging-overview

Diffraction
- ADXV A program to display X-Ray diffraction images
- ***Binoculars*** Surface X-ray diffraction 2D detector data reduction
- EXPGUI is a graphical interface for the GSAS package
- *Dioptas* X-ray diffraction GUI on top of PyFAI
- Fox, 'Free Objects for Crystallography'
- ***gHKL*** is an interface to the hkl library
- *xrayutilities* python library
- *Xraylib* library for X-ray computation
- *Libxy-bin* xylib - utilities for (x,y) powder diffraction
- *Pyfai* Fast Azimuthal Integration scripts (python library) (GPU)
- *PyNX* Python tools for Nano-structures Xtallography (GPU)
- *Spd* Synchrotron image corrections and azimuthal integration (better use pyfai)

Diffraction / small angle scattering
- SASfit Software package SASfit for fitting small-angle scattering curves
- SasView is a Small Angle Scattering Analysis Software Package
- **Foxtrot** SAXS Data treatment

Spectroscopy
- *Artemis/Athena* absorption Xray spectroscopy data analysis
- *ARPYS* python library for ARPES (Angle Resolved PhotoEmission Spectroscopy)
- aXis 2000 - Analysis of X-ray Images and Spectra
- *Fastosh* XAS data treatment (absorption)
- *HyperSpy* multi-dimensional data analysis toolbox
- iFEFFit Interactive XAFS analysis program
- ***NavARP*** Navigation tools for Angle Resolved Photoemission spectroscopy data
- *pymca* PyMca for XRF, Powder diffraction, XAS, FT-IR, Raman, microscopy, fitting ...
- *Mantis Xray* spectro-microscopy and tomography
- *XRSTools* (XRS_raman_extraction, XRS_roiNmaSelection)

Imaging / microscopy
- ADXV A program to display X-Ray diffraction images
- AMIDE's a Medical Image Data Examiner
- *aXis 2000* - Analysis of X-ray Images and Spectra
- CubeView 3D FITS data viewer specialized in spectro-imaging
- ImageJ/Fiji an improved ImageJ with plenty of plugins
- Gwyddion visualization and analysis of data from scanning probe microscopy (SPM)
- imview Image viewing and analysis application
- *Mantis Xray* spectro-microscopy and tomography
- qnifti2dicom – convert 3D medical images to DICOM 2D series
- Relion a CryoEM imaging tool (GPU)
- scikit-image Python 3 modules for image processing
- xmedcon stands for Medical Image Conversion

Volume viewer/modeller
- Geomview is an interactive 3D viewing program (OFF)
- gmsh Three-dimensional finite element mesh generator
- gyotoy utility program uses gyoto framework to compute and display a single geodesic
- Mayavi2 is a scientific data visualizer
- meshlab processing and editing of unstructured 3D triangular meshes
- MRIcron is a cross-platform NIfTI format image viewer.
- MRtrix3/MRview MRI analysis
- ODIN is a framework for magnetic resonance imaging (MRI)
- ParaView data analysis and visualization application (GPU)
- view3dscene is a viewer for many 3D model formats

Volume reconstruction
- *astra-toolbox* modules for octave and python (GPU)
- invesalius 3D medical imaging reconstruction/segmentation software
- ITK-SNAP is a software application used to segment structures in 3D medical images.
- *Mantis Xray* spectro-microscopy and tomography
- *PyHST2* Python High Speed Tomographic reconstruction (GPU)
- ***UFO*** Library for high-performance, GPU-based computing - tools (GPU)

Simulation
- CTSim simulates the process of transmitting X-rays through phantom objects
- GEANT4 simulation of the passage of particles through matter
- ***McXtrace*** general Monte Carlo ray-tracing: X-ray beamlines and experimentssimulation
- ODIN is a framework for magnetic resonance imaging (MRI)
- Pulsar A graphical user interface for the generation and simulation of RF pulses
- *Spectra* a synchrotron radiation calculation code
- Molecular Dynamics / ab-initio / ASE…

Data Viewer
- *edfviewer* Simple EDF file viewer (PyMCA)
- *elementsinfo* - Periodic table with Atomic Constants used by PyMca
- Grace is a WYSIWYG 2D plotting tool (Xmgr/Xmgrace)
- HDFCompass a visual tool to navigate HDF5 files and other resources (local and remote).
- Kst is a fast real-time large-dataset viewing and plotting tool
- LabPlot similar to Origin and Kaleidagrah
- *peakidentifier* - Displays X-ray fluorescence peaks in a given energy range
- *pymca* PyMca XRF, Powder diffraction, XAS, FT-IR, Raman, microscopy, fitting ...
- *silx view* viewer (HDF5) from silx project
- Veusz 2D and 3D plotting
- VisTrails simulations, data exploration and visualization
- ViTables browsing and editing files in both PyTables and HDF5 formats
- Matlab, Julia, R, Jupyter Notebooks...

Legend :
Bold=software we develop/contribute
italic=software we have packaged for Debian

All of this in a Debian SOLEIL Data Analysis VM (QCOW2)

Available for all beam-lines at SOLEIL Synchrotron.

We also provide other environments :
- **AlphaFold (req. GPU)**
- **Windows 10 (with Igor pro, Crysalis, CasaXPS, OPUS, ...)**

DARTS and VISA (PaNOSC/ExPaNDS)

Today, our VM's are being made compatible with OpenStack/VISA (cloud-init).

We shall transfer gradually our VM's into VISA (handled by IT), in order to focus on the scientific applications. The VM scientific contents will remain ours.

	DARTS	OpenStack/VISA
Scalability	yes (kind of) Heterogeneous	**YES** Integrated
Failure strategy	**None**	**YES**
Deployment	Small scale (lab) Easy	Large scale Complex (OpenStack)
Complexity/ Maintenance	Low	**High**

DARTS will probably remain on our development server.

The DEMO effect

Just for today: http://79.93.150.7/qemu-web-desktop (at my home).
And locally: http://localhost/qemu-web-desktop/