

El software científico en el CESGA

Software disponible y apoyo en el despliegue



Jornada de Usuarios CESGA 2023 - 14 de diciembre de 2023

Jorge Fernández Fabeiro - Técnico Superior de Aplicaciones

14/12/2023



XUNTA
DE GALICIA



GOBIERNO
DE ESPAÑA

MINISTERIO
DE CIENCIA
E INNOVACIÓN



CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS



FEDER - FONDO EUROPEO DE
DESENVOLVIMENTO REGIONAL
"Unha maneira de facer Europa"

UNIÓN EUROPEA

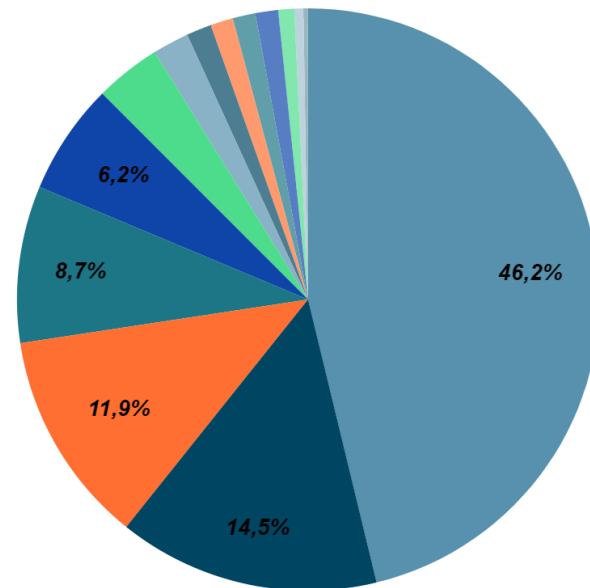
Software disponible

14/12/2023

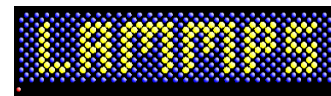


Gestionamos ~800 aplicaciones!

Bioinformática	366
Herramientas de propósito general	115
Visualización y formatos de datos	94
Química y materiales	69
Librerías matemáticas	49
Compiladores e intérpretes	29
Multifísica y CFD	16
Profiling y rendimiento	11
Simulación	10
MPI	10
Machine Learning	10
Quantum Computing	7
Análisis numérico y estadístico	4
Gestión de software	2

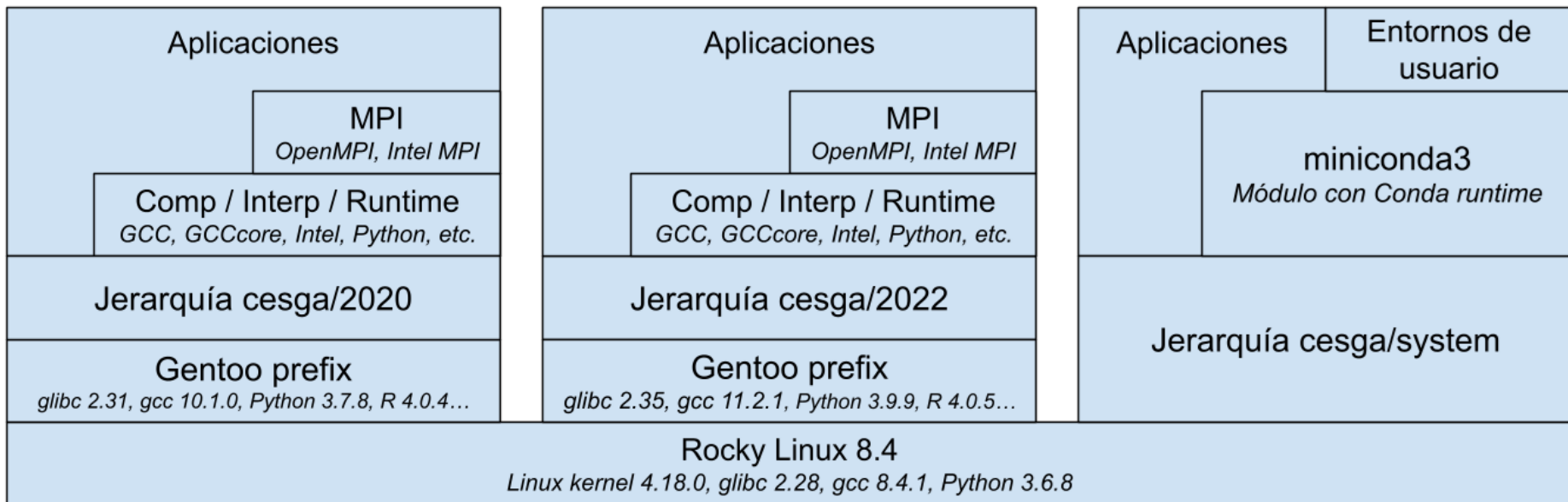


Aplicaciones más usadas* en 2023

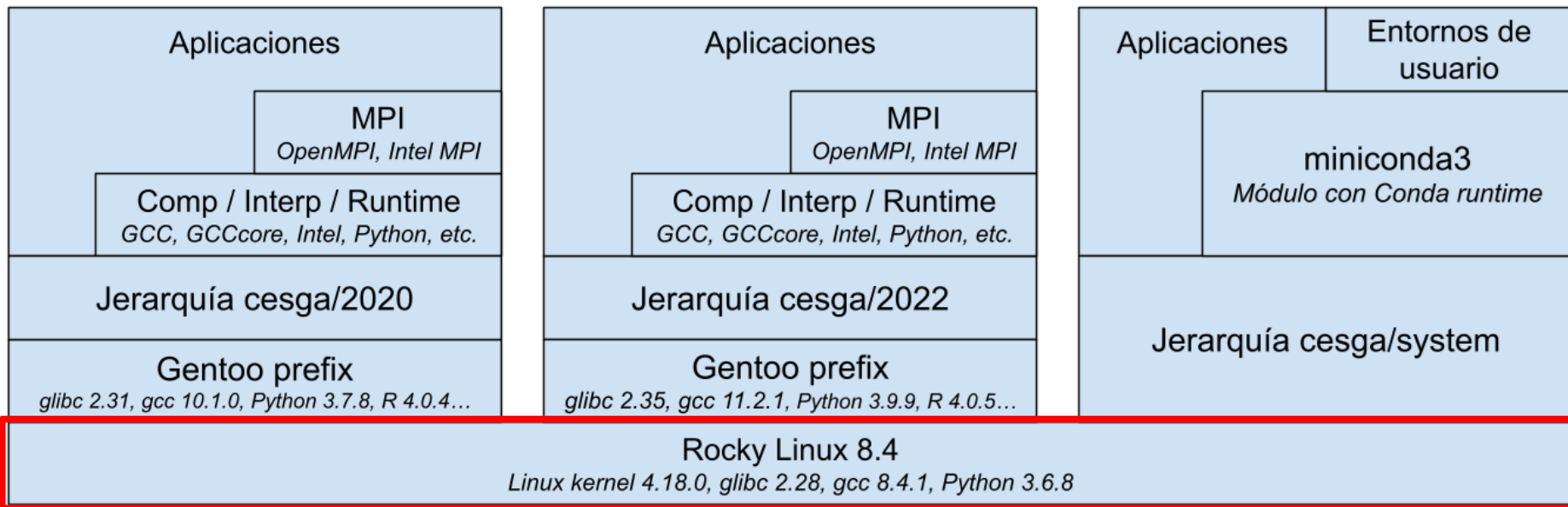


(h. uso, *orden aprox.)

Organización general

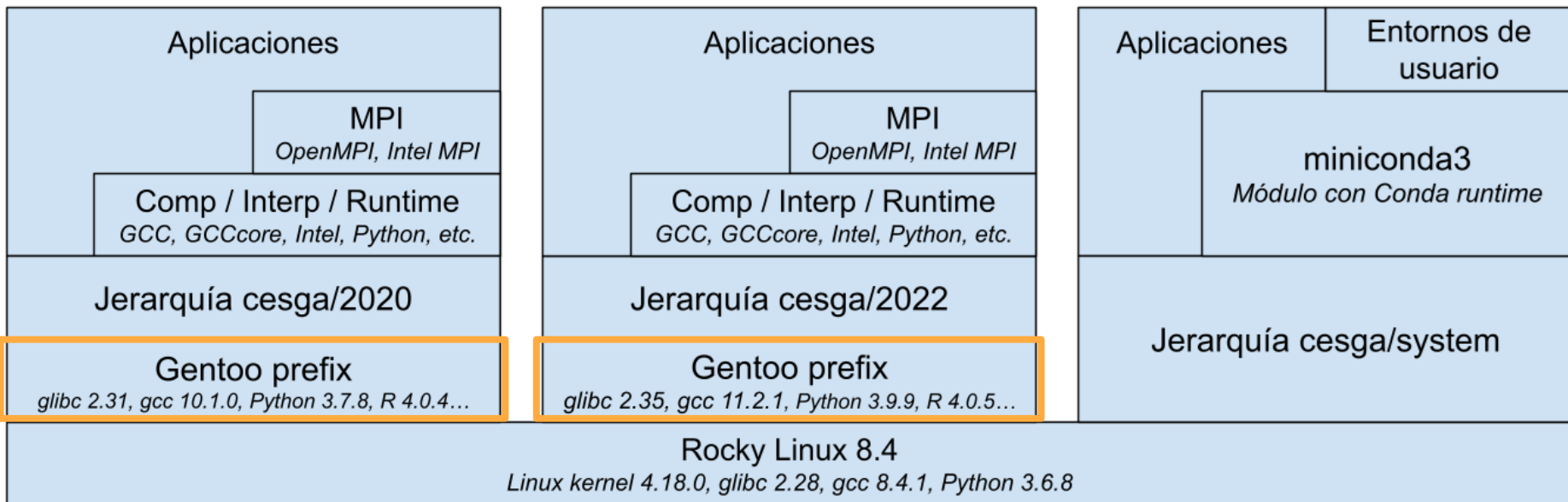


Organización general



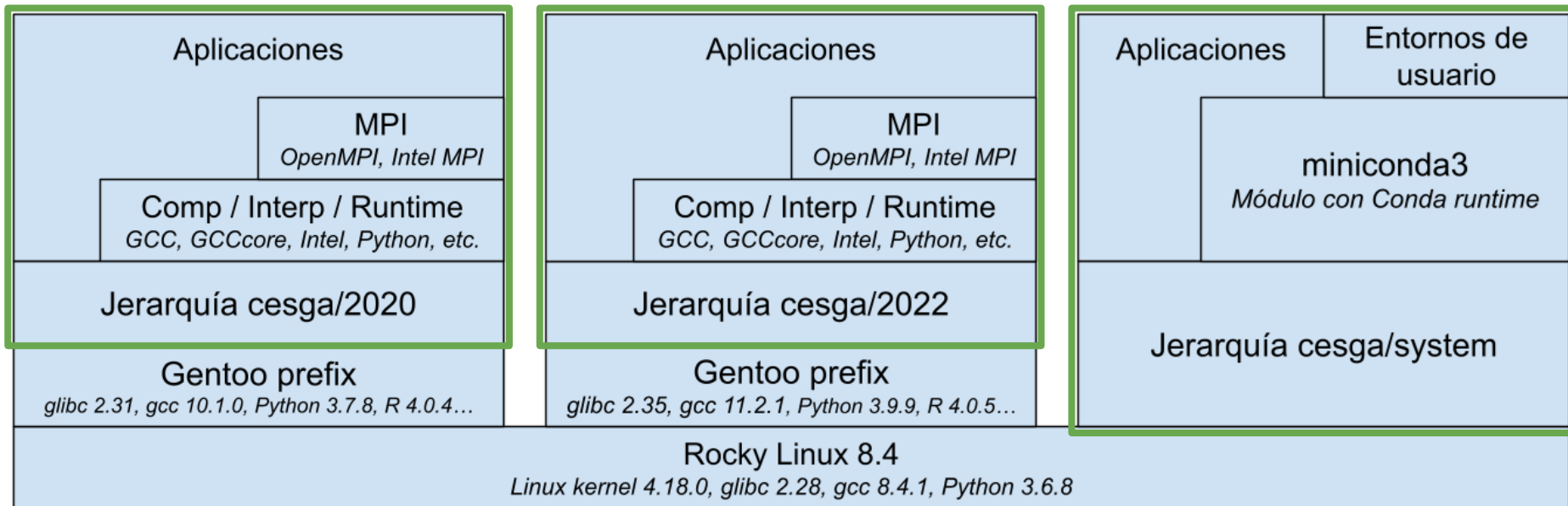
**S.O. de las máquinas del FT3: cerrado y estable
Mantenido por el departamento de Sistemas**

Organización general



Base de SW completa, flexibilidad para instalar paquetes
Mantenidos por el departamento de Aplicaciones

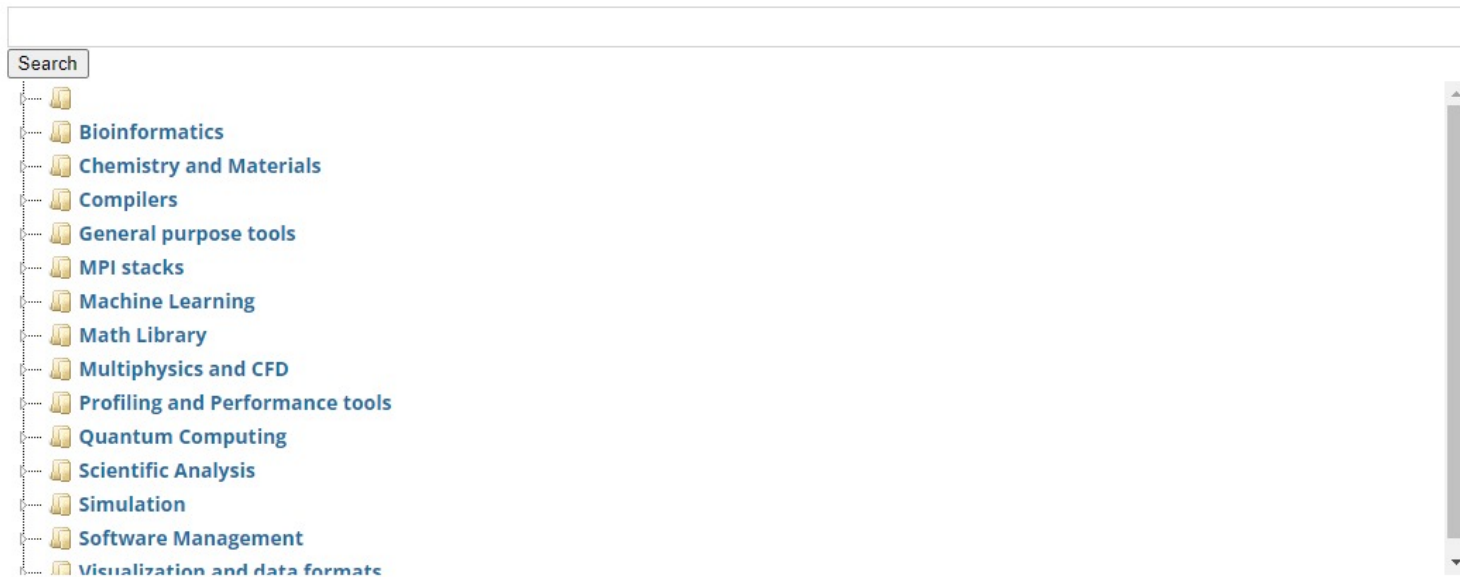
Organización general



Módulos con aplicaciones
Mantenidos por el departamento de Aplicaciones

Cómo saber si una aplicación está disponible

Buscador de módulos en <https://www.cesga.es/infraestructuras/aplicaciones/>



Cómo saber si una aplicación está disponible

Con acceso al FT3:

¿La ofrece algún paquete del Gentoo prefix de cesga/202X? ¿Me sirve esa versión?

```
[jfernandez@login211-1 ~]$ equery l samtools
* Searching for samtools ...
[IP-] [ ] sci-biology/samtools-1.12:0
```

```
equery l <paquete>
```

Si no está en Gentoo o la versión no me sirve, ¿está disponible como módulo bajo esa jerarquía?

```
[jfernandez@login211-1 ~]$ module spider samtools
```

```
module spider <cadena>
```

```
samtools:
```

```
Description:
```

```
SAM Tools provide various utilities for manipulating alignments in the SAM format, including sorting, merging, indexing and generating alignments in a per-position format. --
cesga/2020 samtools/0.1.20 : Core
```

```
Versions:
```

```
samtools/0.1.16
samtools/0.1.20
samtools/1.9
samtools/1.14
```

```
For detailed information about a specific "samtools" package (including how to load the modules) use the module's full name.
```

```
Note that names that have a trailing (E) are extensions provided by other modules.
```

```
For example:
```

```
$ module spider samtools/1.14
```

Cómo saber si una aplicación está disponible

Con acceso al FT3:

Detalles del módulo concreto que interesa (descripción y cadenas de carga):

```
[jfernandez@login211-1 ~]$ module spider samtools/1.9
```

```
-----  
samtools: samtools/1.9  
-----
```

Description:

```
SAM Tools provide various utilities for manipulating alignments in the SAM format, including sorting, merging, indexing and generating alignments in a per-position format. -- cesga/2020 samtools/1.9: Compiler: Requires gcccore/system
```

Properties:

```
Bioinformatics (key: area_bioinformatics)
```

```
You will need to load all module(s) on any one of the lines below before the "samtools/1.9" module is available to load.
```

```
cesga/2020  
cesga/2020 gcccore/system
```

```
module load cesga/2020 samtools/1.9  
module load cesga/2020 gcccore/system samtools/1.9
```

Cómo saber si una aplicación está disponible

Con acceso al FT3:

Detalles del módulo concreto que interesa
(ayuda y ejemplos de uso):

¡Cuidado! “Usage” indica un ejemplo de cadena de carga, si el módulo admite varios puede no ser el deseado.

```
Help:
Description
SAM Tools provide various utilities for manipulating alignments in the SAM format,
including sorting, merging, indexing and generating alignments in a per-position format.

Usage
INTERACTIVE USE:
module load cesga/2020 gcccore/system samtools/1.9
samtools <command> [options]

QUEUE SYSTEM:
-SEQUENTIAL:
Recommended script (run.sh):

#!/bin/bash
#SBATCH -t 01:00:00 # execution time
module load cesga/2020 gcccore/system samtools/1.9
samtools <command> [options]

sbatch run.sh

SPECIFIC INFO:
Command line help: samtools --help

More information
- Homepage: http://www.htslib.org/
- Documentation:
  - http://www.htslib.org/doc/
```

Cómo saber si una aplicación está disponible

Casos especiales: Aplicaciones con soporte MPI

Comprobar que algún módulo de MPI está en las posibles cadenas de carga

```
[jfernandez@login209-18 /opt/cesga/2020/repo]$ module spider openfoam/v2106
-----
openfoam: openfoam/v2106
-----
Description:
  OpenFOAM is a free, open source CFD software package. OpenFOAM has an
  turbulence and heat transfer, to solid dynamics and electromagnetics.

Properties:
  Multiphysics and CFD (k

You will need to load all
  cesga/2020 gcc/system openmpi/4.0.5
  cesga/2020 gcc/system openmpi/4.0.5_ft3
  cesga/2020 gcc/system openmpi/4.1.4_ft3
  cesga/2020 intel/2021.3.0 impi/2021.3.0
```

```
module load cesga/2020 gcc/system openmpi/4.1.4_ft3 openfoam/v2106
module load cesga/2020 intel/2021.3.0 impi/2021.3.0 openfoam/v2106
```

Cómo saber si una aplicación está disponible

Casos especiales: Intérpretes Python 3 y paquetes

Hay varios intérpretes de Python 3 en cesga/2020

Python 3.7.8 a través del Gentoo prefix

```
[jfernandez@login211-1 ~]$ equery l python-3.7*  
* Searching for python-3.7* ...  
[IP-] [ ] dev-lang/python-3.7.8-r2:3.7/3.7m
```

Python 3.6.12 y 3.9.9 como módulos

```
[jfernandez@login211-1 ~]$ module list  
  
Currently Loaded Modules:  
 1) cesga/2020 (S)  
  
[jfernandez@login211-1 ~]$ module avail python/3  
  
----- Core -----  
python/3.6.12 (Comp)    python/3.9.9 (Comp,D)
```

Cómo saber si una aplicación está disponible

Casos especiales: Intérpretes Python 3 y aplicaciones/paquetes

Combinación adecuada de intérprete y aplicaciones/paquetes

Si procede del Gentoo o de un módulo sin sufijo, el intérprete adecuado es el Python 3.7.8 por defecto:

```
[jfernandez@login211-1 ~]$ equery l python-3*
* Searching for python-3* ...
[IP-] [ ] dev-lang/python-3.7.8-r2:3.7/3.7m
[jfernandez@login211-1 ~]$ equery l pandas
* Searching for pandas ...
[IP-] [ ] dev-python/pandas-1.0.5:0
[jfernandez@login211-1 ~]$ python
Python 3.7.8 (default, Nov 27 2020, 12:00:52)
[GCC 10.1.0] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> import pandas as pd
>>> pd.__version__
'1.0.5'
>>> pd.__path__
['/mnt/netapp1/0ptcesga_FT2_RHEL7/2020/gentoo/22072020/usr/lib/python3.7/site-packages/pandas']
```

Cómo saber si una aplicación está disponible

Casos especiales: Intérpretes Python 3 y aplicaciones/paquetes

Combinación adecuada de intérprete y aplicaciones/paquetes

Si procede de un módulo con sufijo `-python-X.X.X`, se carga automáticamente el intérprete adecuado:

```
[jfernandez@login211-1 ~]$ module spider pandas
-----
pandas:
-----
Description:
pandas is an open source, BSD-licensed li
cesga/2020 pandas/1.3.5-python-3.9.9 : Co

Versions:
pandas/1.1.5-python-3.6.12
pandas/1.3.2
pandas/1.3.4-python-3.9.9
pandas/1.3.5-python-3.9.9
```

```
[jfernandez@login211-1 ~]$ module load pandas/1.3.5-python-3.9.9
python/3.9.9 loaded
imkl/2021.3.0 loaded
numpy/1.23.5-python-3.9.9 loaded
pandas/1.3.5-python-3.9.9 loaded
```

¡No es posible combinar módulos que necesiten intérpretes distintos, ¡se autoexcluyen!

Cómo saber si una aplicación está disponible

Casos especiales: Aplicaciones ofrecidas como entornos Conda

Utilizados para instalar determinadas aplicaciones con **elevada complejidad de dependencias**

Son cerrados, los usuarios no pueden instalar más paquetes en ellos

Uso aislado, salvo situaciones concretas no se recomienda / no es posible combinarlos con otros módulos

```
[jfernandez@login211-1 ~]$ module spider pytorch/2.0.0-cuda
-----
pytorch: pytorch/2.0.0-cuda
-----
Description:
  Tensors and Dynamic neural networks
  Core

Properties:
  Machine Learning (key: area_machinele

You will need to load all module(s) on any one of the l

cesga/system

[jfernandez@login211-1 ~]$ module load cesga/system pytorch/2.0.0-cuda
cesga/system loaded
miniconda3/22.11.1-1 loaded
pytorch/2.0.0-cuda loaded

[jfernandez@login211-1 ~]$ conda info
active environment : /opt/cesga/system/software/Core/pytorch/2.0.0-cuda
```

Cómo saber si una aplicación está disponible

Casos especiales: R

R/4.0.4 es ofrecido por el Gentoo prefix de `cesga/2020` y puede enriquecerse cargando `bioconductor/3.12`

```
[jfernandez@login211-1 ~]$ equery l R
* Searching for R ...
[IP-] [ ] dev-lang/R-4.0.4:0
```

\$ R # Accesible directamente, sin módulo

```
[jfernandez@login211-1 ~]$ module spider bioconductor/3.12
-----
bioconductor: bioconductor/3.12
-----
Description:
Bioconductor provides tools for the analysis and coprehension of
high-throughput genomic data. -- cesga/2020 bioconductor/3.12 : Core

Properties:
Bioinformatics (key: area_bioinformatics)

You will need to load all module(
"bioconductor/3.12" module is availa

cesga/2020
```

module load cesga/2020 bioconductor/3.12
\$ R # Verá los paquetes de bioconductor!

Cómo saber si una aplicación está disponible

Casos especiales: R

R/4.2.2 es un entorno Conda disponible como módulo de **cesga/system**

```
[jfernandez@login211-1 ~]$ module spider R/4.2.2
```

```
-----  
R: R/4.2.2  
-----
```

Description:

R is GNU S, a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modelling, statistical tests, time series analysis, classification, clustering, etc.

Properties:

Bioinformatics (key: area_bioinformatics)

You will need to load all module(s) on any one of the lines below before the "R/4.2.2" module is available to load.

```
cesga/system
```

```
module load cesga/system R/4.2.2
```

Apoyo en el despliegue

Caso general

Solicitud de instalación de aplicación no disponible en el FT3

Solicitud

Estudio

Respuesta

Apertura de ticket al servicio de soporte de aplicaciones:

- helpdesk_aplicaciones@cesga.es
- Portal Usuarios

Información necesaria:

- Nombre y descripción de la app
- Medio de acceso al software
- Licencia de uso:
 - Privada? Uso comercial?

Los técnicos del departamento estudiamos las posibles opciones para desplegar la aplicación:

- Módulo vs. manual
- Público vs. restringido
- Entorno Conda
- Contenedor Singularity
- Complementos paquetes R
- Compilación manual.
- etc.

Una vez desplegada lo notificamos y damos unas pautas mínimas para comenzar a usarla:

- Dónde está
- Cómo se carga/accede
- Quién podrá usarla
- Instrucciones adicionales
- Advertencias/consejos
- etc.

Otros escenarios

Prueba de aplicaciones en la cuenta de usuario

No es necesario solicitar la instalación para probar aplicaciones de forma autónoma en el FT3

Prestamos asistencia en las diferentes fases del proceso:

- Partes que exigen permisos de administración
 - Instalación de paquetes de sistema (apt, yum, dnf, etc.) y/o reemplazo por módulos cargables
 - Configuración de rutas de instalación (problemas de permisos con make install)
- Creación de entornos propios (**modificables por el usuario!**)
 - Entornos virtuales de Python (combinables con módulos de cesga/202X, soporte para MPI, etc.)
 - Entornos Conda para dependencias complejas (soporte GPU)
- Construcción de contenedores Singularity (repositorios, ficheros de especificación, conversión Docker)
- etc.

Otros escenarios

Traslado de ejecuciones o código propio al FT3

¿Cuentas con alguna simulación, cálculo, pipeline, código, etc. ya preparado en una máquina local?

¿Es muy lento o no dispones de recursos suficientes para el tamaño real del problema?

Algunos ejemplos de asistencia que prestamos:

- Refactorización de código multi-hilo/multi-proceso/multi-nodo/multi-GPU:
 - ML: distribución de conjuntos de datos, entrenamientos y/o inferencia
 - Problemas vergonzosamente paralelos con GNU parallel
 - Paralelización multinivel MPI + OpenMP
- Adaptación de parámetros de entrada (hilos, procesos, memoria...) y envío de scripts a colas
- Composición del entorno de ejecución más apropiado: standalone, virtualenvs, Conda, Singularity, etc.

Otros escenarios

Composición de entornos Conda de usuario





- Creación e instalación libre
- Algunas pautas básicas de uso:
 - Miniconda3 de cesga/system: `module load cesga/system miniconda3`
 - Configurar Conda para guardar entornos y paquetes en \$STORE (ver Guía FT3 o consultarnos)
 - `mamba create -n <entorno> -c <canal> python=3.X <paquetes>`
 - `conda activate <entorno>`
 - `mamba install -c <canal> <paquete>`
- Paquetes para soporte CUDA habitualmente como dependencias de paquetes principales
- MPI posible pero no recomendado (paquetes con implementaciones propias... ¿optimización, SLURM?)

Otros escenarios

Instalación de paquetes R

- De entrada, posible en la carpeta personal por defecto del usuario
- La versión es importante, compatibilidad R 4.0 y 4.2 poco habitual
- Los paquetes se compilan al ser instalados, **¡pueden faltar librerías de sistema!**
 - R 4.0.4 de Gentoo o R 4.2.2 de cesga/system: consultar con Aplicaciones
- **Alternativa autónoma en cuenta de usuario**
 - Crear un entorno Conda propio con vuestra propia versión de R y paquetes
 - Instalación de paquetes desde R y gestión de dependencias de sistema como paquetes Conda
 - Canales Conda ofrecen implementaciones completas (dependencias de sistema + paquetes)

Portal de Usuarios CESGA -> Información -> Enlaces

- Sección Computación del web del CESGA: <https://www.cesga.es/infraestructuras/computacion>
- Aplicaciones instaladas en los servidores de cálculo: <https://www.cesga.es/infraestructuras/aplicaciones/>
- Guía de uso del FinisTerra III: <https://cesga-docs.gitlab.io/ft3-user-guide/index.html>
- **Talleres del FinisTerra III:**
 - Taller de introducción:
 1. Presentación: 
 - Taller avanzado:
 1. Presentación de la sesión 1: 
 2. Presentación de la sesión 2: 
 3. Vídeos: https://youtube.com/playlist?list=PLvAJrckik_UMoHPWYYz-SqY4asXP2zvMN
- Transferencia de datos: https://bigdata.cesga.es/dtn-user-guide/CESGA_DTN_User_Guide.pdf
- Guía de uso del Cloud: 
- Solicitud de almacenamiento: <https://altausuarios.cesga.es/solic/almac>
- Solicitud de recursos especiales: <https://altausuarios.cesga.es/solic/recesp>

Gracias por vuestra atención

Correo individual: jfernandez@cesga.es

Consultas generales: aplicaciones@cesga.es

14/12/2023



XUNTA
DE GALICIA



GOBIERNO
DE ESPAÑA

MINISTERIO
DE CIENCIA
E INNOVACIÓN



CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS



FEDER - FONDO EUROPEO DE
DESENVOLVIMIENTO REGIONAL
"Unha maneira de facer Europa"