HPC Now!

Integrating R with HPC

Success stories form pharmaceutical research

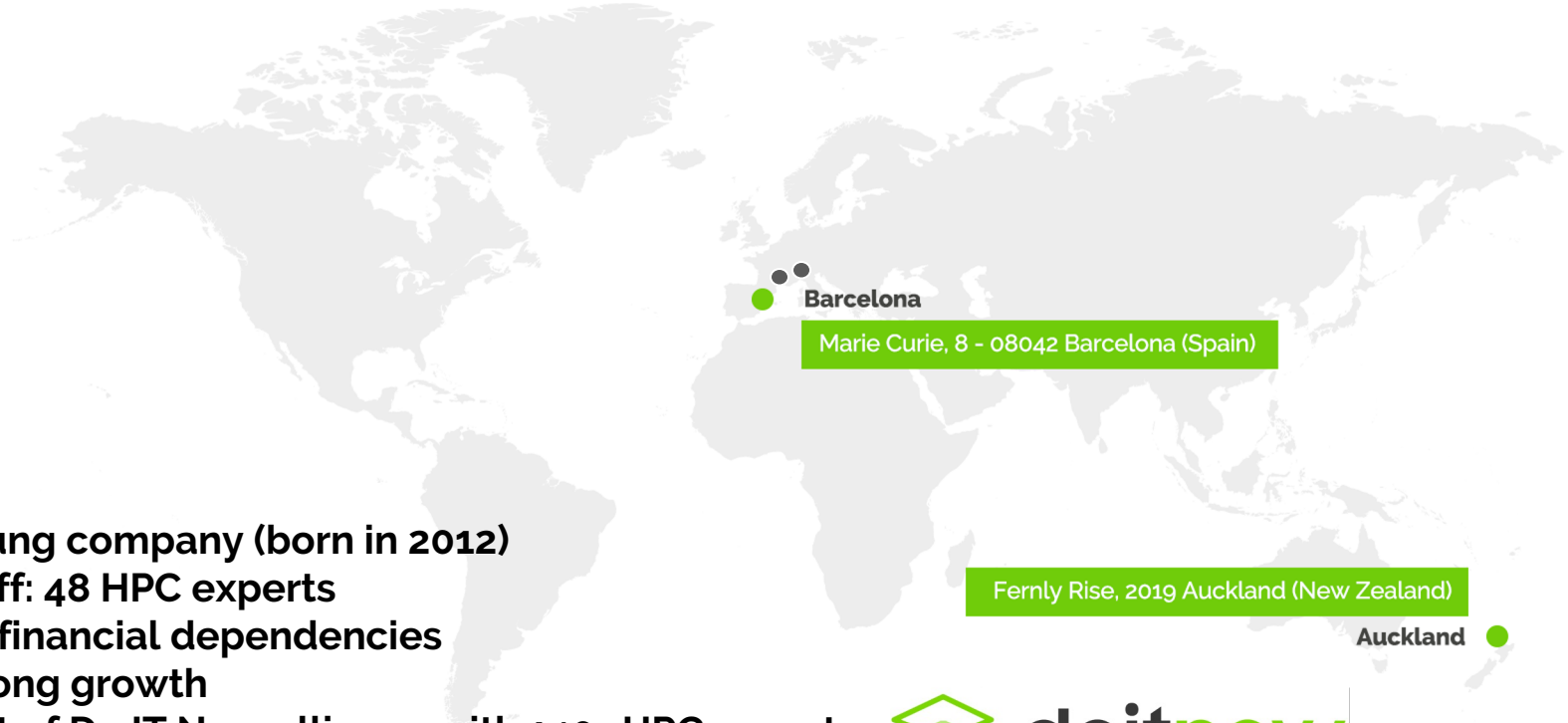
Guillem de Valles Ibáñez

8/05/2024

www.hpcnow.com

Overview - HPCNow

Contact



- Young company (born in 2012)
- Staff: 48 HPC experts
- No financial dependencies
- Strong growth
- Part of Do IT Now alliance with 140+ HPC experts
- Hiring +20 more in Q2

doitnow HPC Services

Overview - HPCNow

- **Customers**

Pfizer
Roche
SIEMENS Gamesa RENEWABLE ENERGY
arianeGROUP
Applus+ IDIADA

KAUST
BSC Barcelona Supercomputing Center Centro Nacional de Supercomputación
ALBA
EMBL-EBI

CSIRO
pawsey
Fred Hutch Cancer Center
GNS SCIENCE TE PŪ AO

and many more...

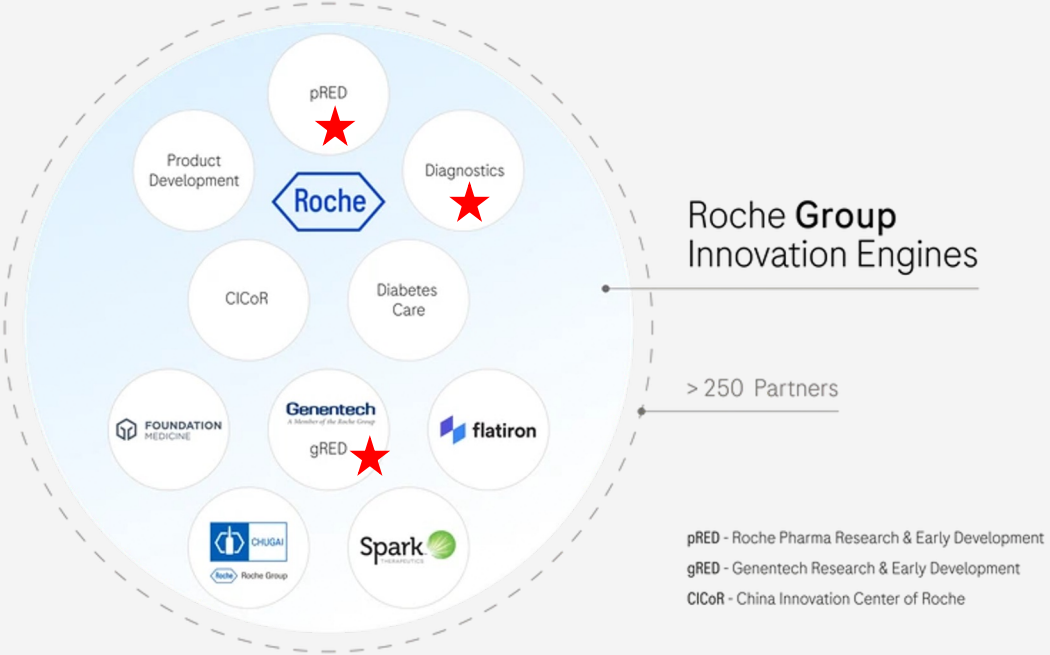
Overview - Roche

F. Hoffmann-La Roche AG

5th largest pharmaceutical by revenue

Swiss multinational with headquarters in Basel

- Founded in 1896
- Revenue of 122.05 billion NZD (2022)
- 103,613 employes (2022)
- Acquired 50 companies
- 2 major divisions: Pharmaceutical and diagnostics
- Focus on health care applications (targeting physicians, hospitals, clinics and consumers) and pharmaceutical and biotechnology research



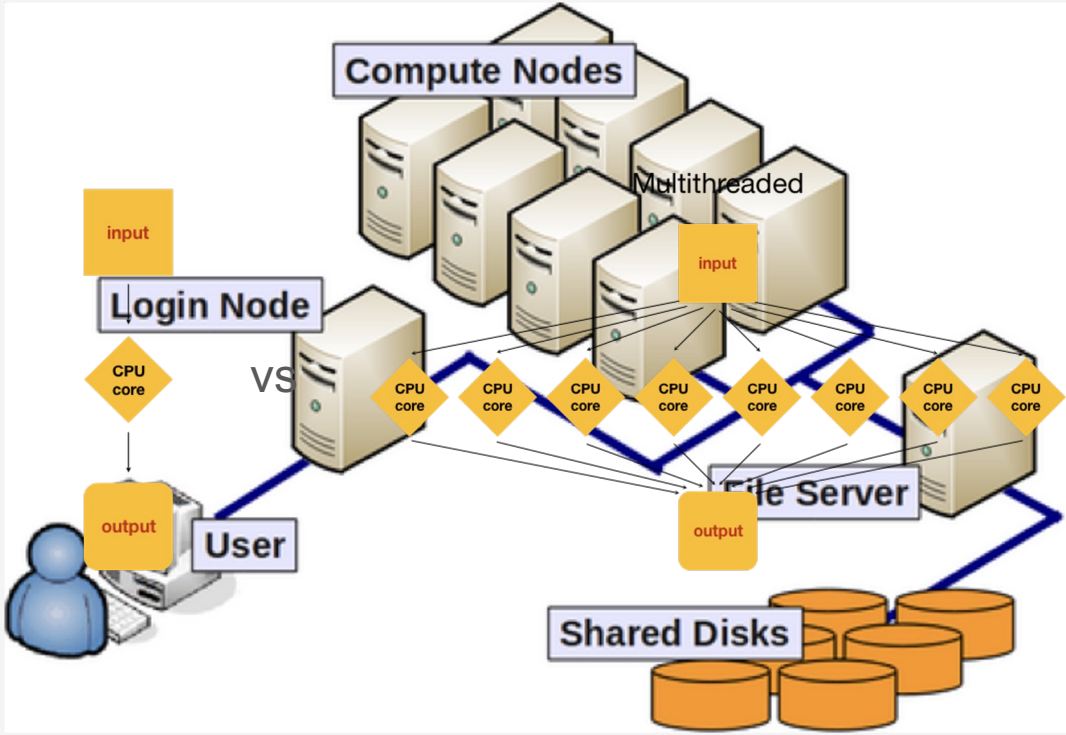
Overview - HPC

High Performance Computing clusters

High Performance Computing

Use of computer clusters to solve advanced computation problems

- 100s of users
- 100s of apps
- 1000s of cores for processing
- 100s of Tb of storage
- 100s of Gb of memory (per node)
- 100s of GPUs
- Job management and monitoring
- Interactive applications (Jupyter notebooks, RStudio,...)



Overview - HPC

Top 500 (November 2023)

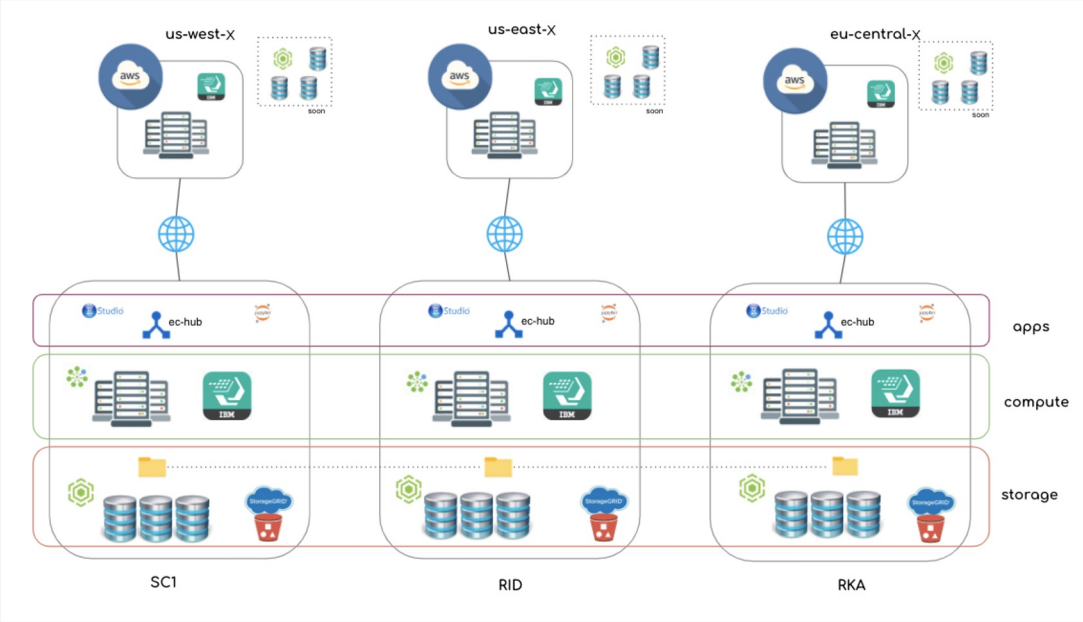
Rank (previous)	Rmax Rpeak (PetaFLOPS)	Name	CPU cores	Accelerator (e.g. GPU) cores	Total Cores (CPUs + Accelerators)	Manufacturer	Site country	Year
1	1,194.00 1,679.82	Frontier	561,664 (8,776 × 64-core)	36,992 × 220 AMD Instinct MI250X	8,699,904	HPE	Oak Ridge National Laboratory United States	2022
2	585.34 1,059.33	Aurora	565,656 (10,878 × 52-core)	32,634 × 128 Intel Max 1550	4,742,808	HPE	Argonne National Laboratory United States	2023
3	561.20 846.84	Eagle	93,600 (1,950 × 48-core)	7,800 × 132 Nvidia Hopper H100	1,123,200	Microsoft	Microsoft United States	2023
4	442.010 537.212	Fugaku	7,630,848 (158,976 × 48-core)	-	7,630,848	Fujitsu	RIKEN Center for Computational Science Japan	2020
5	309.10 428.70	LUMI	186,624 (2,916 × 64-core)	11,664 × 220 AMD Instinct MI250X	2,752,704	HPE	EuroHPC JU European Union, Kajaani, Finland	2022
6	238.70 304.47	Leonardo	110,592 (3,456 × 32-core)	15,872 × 108 Nvidia Ampere A100	1,824,768	Atos	EuroHPC JU European Union, Bologna, Italy	2023
7	148.600 200.795	Summit	202,752 (9,216 × 22-core)	27,648 × 80 Nvidia Tesla V100	2,414,592	IBM	Oak Ridge National Laboratory United States	2018
8	138.20 265.57	MareNostrum 5 ACC	89,600 (2,240 × 40-core)	4,480 × 132 Nvidia Hopper H100	680,960	BullSequana	**EuroHPC JU European Union, Barcelona, Spain**	2023
9	121.40 188.65	Eos NVIDIA DGX SuperPOD	46,592 (832 × 56-core)	3,328 × 132 Nvidia Hopper H100	485,888	Nvidia	Nvidia United States	2023
10	94.640 125.712	Sierra	190,080 (8,640 × 22-core)	17,280 × 80 Nvidia Tesla V100	1,572,480	IBM	Lawrence Livermore National Laboratory United States	2018

Overview - sHPC

Roche's shared High Performance Computing clusters

The Shared-HPC Service (sHPC) provides an ecosystem for High-Performance Compute (HPC) and High-Throughput Compute (HTC) workloads with cloud-bursting capabilities via an Intelligent Scheduler that coordinates the execution of containerized and non-containerized workloads, close to the data, maximizing productivity and resource utilization.

We also have an HPC called Rosalind, exclusive for gRED users and to be decommissioned at the end of 2024



Software provisioning - RoCS

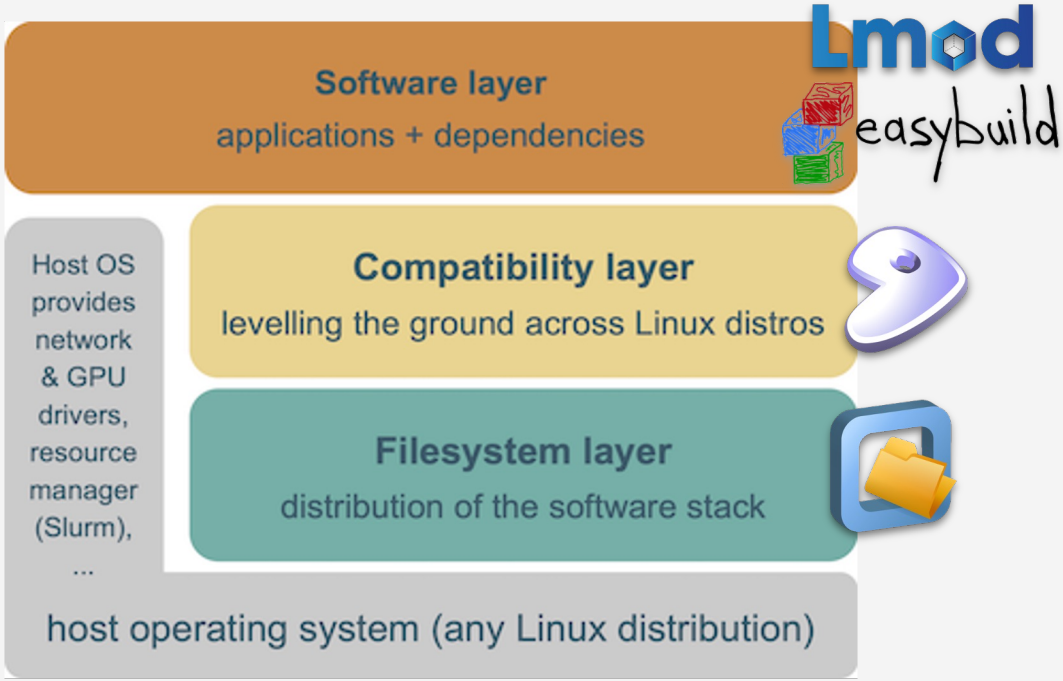
Roche Computing Stack

RoCS appstack

Provides a unified, portable, modern scientific applications environment

- Community based
- Inspired by Compute Canada, the European Environment for Scientific software Installations (EESSI) and CERN models

RoCS Architecture:



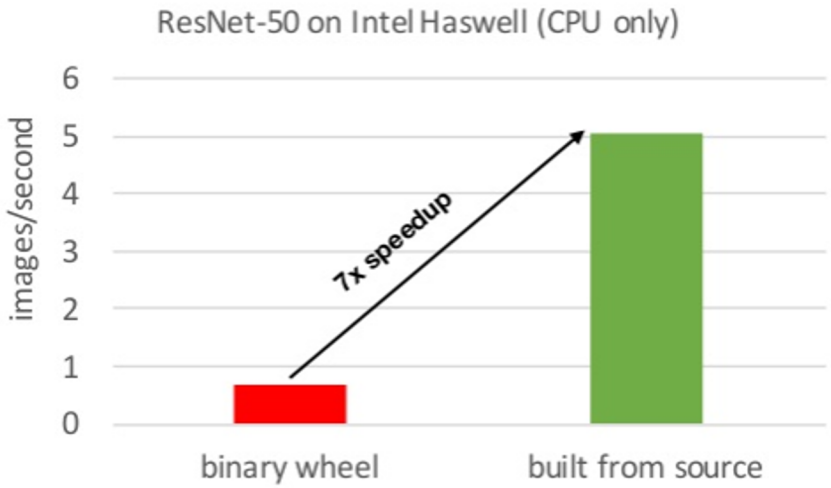
Software provisioning - why build from source?

Speed!

Micro architecture awareness

Building packages from source in an HPC can significantly increase efficiency and speed

- Binaries from repositories have no microarchitecture optimisation for portability
- Precompiled binaries from old architectures might work in new ones with a performance penalty
- Precompiled binaries from new architectures will not work in older ones due to lack of hardware instructions



Provisioning R

4 ways to provision R for our users

Self provisioning

Users can install the R packages they need in their home or scratch folders as they would in their computers

Modules

Web portal

Containers

Provisioning R

4 ways to provision R for our users

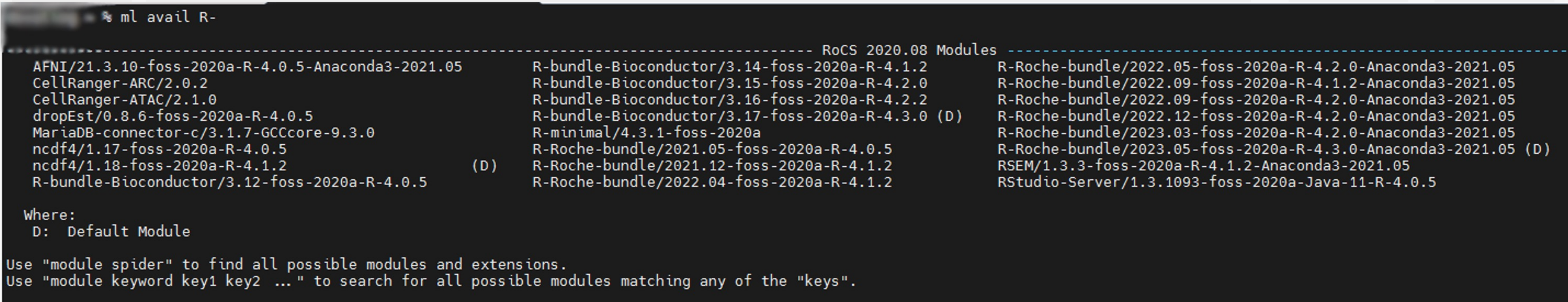
Self provisioning

Modules

The main way to provision apps to our users in the sHPC clusters is through a Lua based module system. Lmod is an implementation of Environment Modules through modulefiles, which contain all the necessary information to allow a user to run a particular application or set of libraries

Web portal

Containers



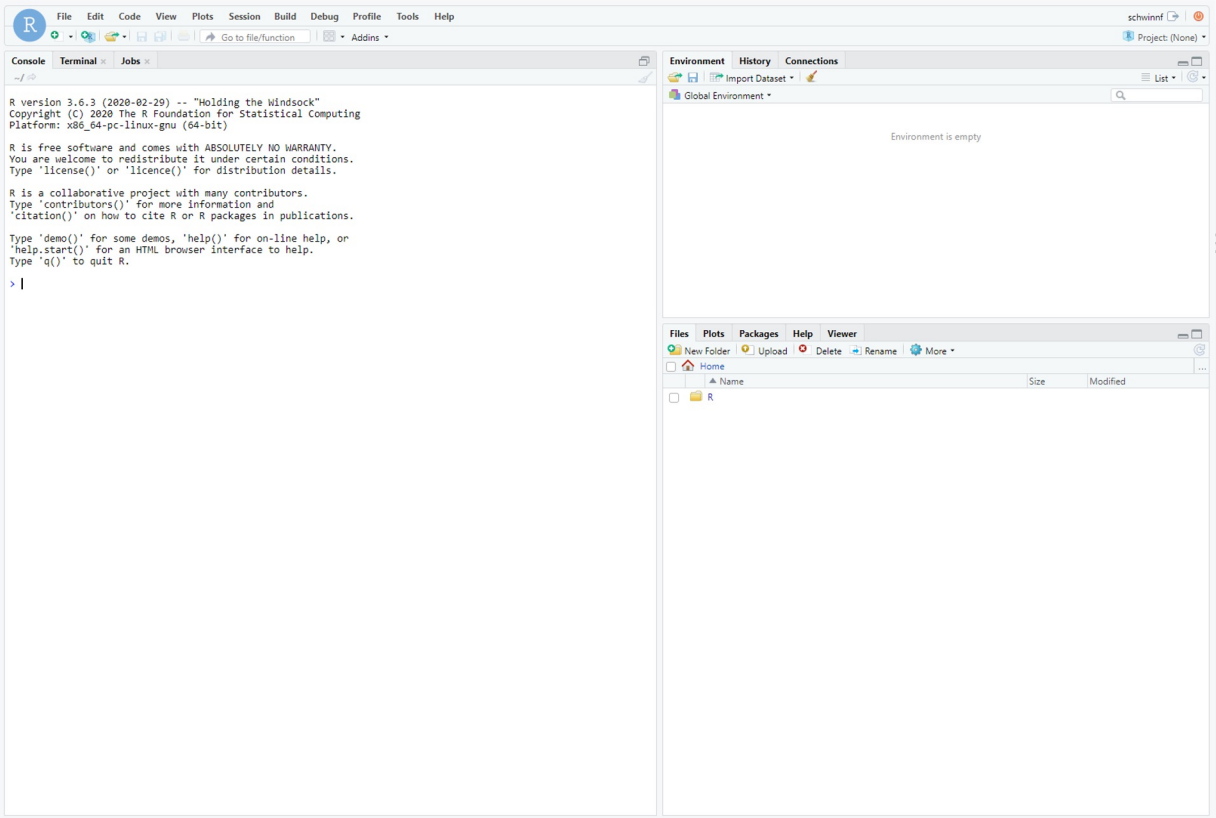
```
[redacted] % ml avail R-

-------------------------------------------------------------------------------------------- RoCS 2020.08 Modules ---------------------------------------------------------------
   AFNI/21.3.10-foss-2020a-R-4.0.5-Anaconda3-2021.05        R-bundle-Bioconductor/3.14-foss-2020a-R-4.1.2        R-Roche-bundle/2022.05-foss-2020a-R-4.2.0-Anaconda3-2021.05
   CellRanger-ARC/2.0.2                                     R-bundle-Bioconductor/3.15-foss-2020a-R-4.2.0        R-Roche-bundle/2022.09-foss-2020a-R-4.1.2-Anaconda3-2021.05
   CellRanger-ATAC/2.1.0                                    R-bundle-Bioconductor/3.16-foss-2020a-R-4.2.2        R-Roche-bundle/2022.09-foss-2020a-R-4.2.0-Anaconda3-2021.05
   dropEst/0.8.6-foss-2020a-R-4.0.5                         R-bundle-Bioconductor/3.17-foss-2020a-R-4.3.0 (D)    R-Roche-bundle/2022.12-foss-2020a-R-4.2.0-Anaconda3-2021.05
   MariaDB-connector-c/3.1.7-GCCcore-9.3.0                  R-minimal/4.3.1-foss-2020a                           R-Roche-bundle/2023.03-foss-2020a-R-4.2.0-Anaconda3-2021.05
   ncdf4/1.17-foss-2020a-R-4.0.5                            R-Roche-bundle/2021.05-foss-2020a-R-4.0.5            R-Roche-bundle/2023.05-foss-2020a-R-4.3.0-Anaconda3-2021.05 (D)
   ncdf4/1.18-foss-2020a-R-4.1.2                     (D)    R-Roche-bundle/2021.12-foss-2020a-R-4.1.2            RSEM/1.3.3-foss-2020a-R-4.1.2-Anaconda3-2021.05
   R-bundle-Bioconductor/3.12-foss-2020a-R-4.0.5            R-Roche-bundle/2022.04-foss-2020a-R-4.1.2            RStudio-Server/1.3.1093-foss-2020a-Java-11-R-4.0.5

  Where:
   D:  Default Module

Use "module spider" to find all possible modules and extensions.
Use "module keyword key1 key2 ..." to search for all possible modules matching any of the "keys".
```

Provisioning R

4 ways to provision R for our users

Self provisioning

Modules



Containers

Provisioning R

4 ways to provision R for our users

Self provisioning

Modules

Web portal

Containers

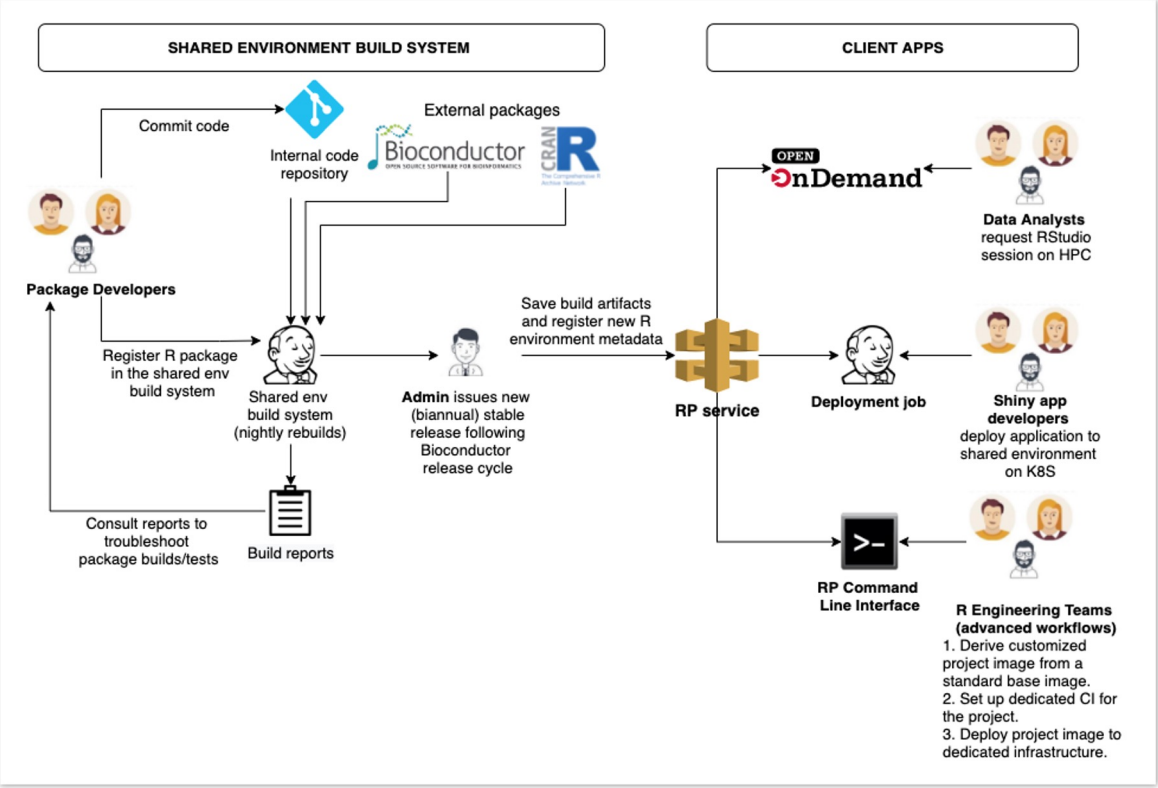
In our sHPC we use Singularity as our container platform, which can also build docker containers. They can be accessed through the RStudio / Posit workbench IDE via OOD or through the RStudio CEDAR module (as in the previous slide)

Provisioning R

Computational Environments for Data Analysis and Research (CEDAR)

CEDAR

- The shared environment build system is responsible for building and testing all packages, which constitute an R environment.
- Stable environments are typically released on a biannual basis and provide versioned images (Docker, Singularity) and their corresponding R package libraries.
- Each release (metadata) is registered using the RP Service REST API.



Provisioning R

Genentech R Archive Network (GRAN)

GRAN

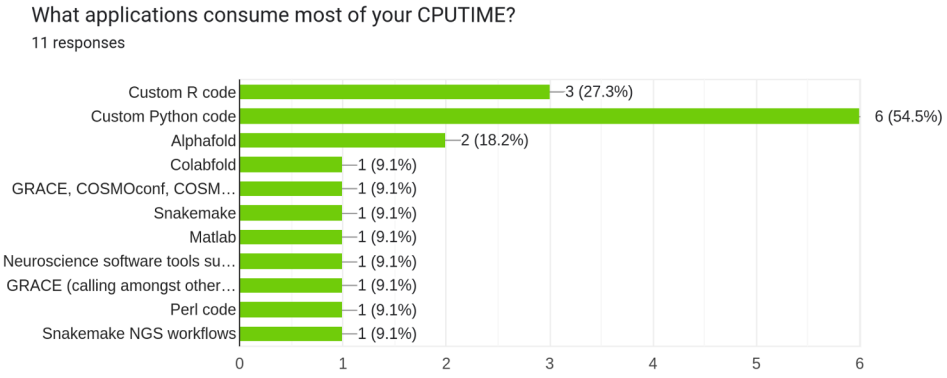
Genentech's internal R package build system and repository, similar to CRAN or Bioconductor.
It provides:

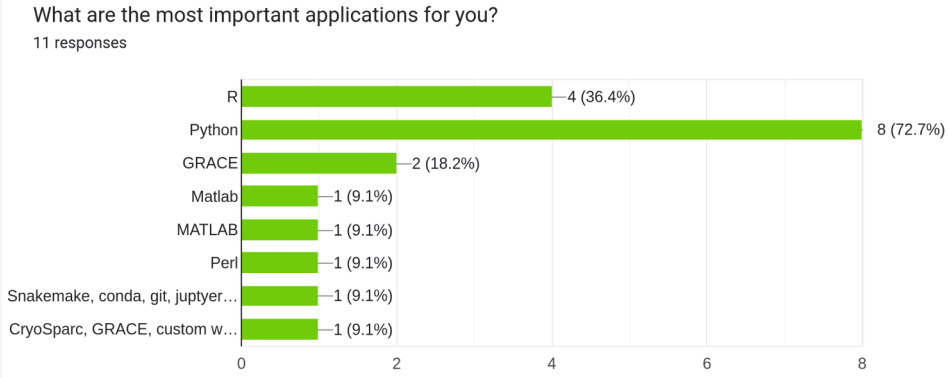
- Continuous, incremental building and testing of internally developed R packages (packages with new versions are built nightly)
- Automatic availability of packages for analysts who work on sHPC
- Live R package repositories and archives for analysts working on other HPCs or computers

Provisioning R

Use of R within Roche

Results from recent user survey





Provisioning R

Use of R within Roche

Posit connect

Shiny server pro to host our shiny apps

- Integration with other apps (snakemake, nextflow, …)
- Optional authentication security (access only to members of a specific UNIX group)
- Application code stored and maintained in GitHub repositories

Thank you

Contact us:
info@hpcnow.com
guillem.devalles@hpcnow.com