



1

L'INFRASTRUTTURA RECAS/IBISCO

Giacinto DONVITO
INFN-Bari



AGGIORNAMENTO SULLO STATO DELLE RISORSE NEI SITI PON-IBISCO/RECAS

On behalf of IBISCO project



PON RICERCA E INNOVAZIONE 2014-2020



Finalità dell'Avviso

Potenziamento delle infrastrutture di ricerca individuate come prioritarie nel PNIR 2014-2020. IR funzionali all'implementazione di progetti rispondenti a ambiti ESFRI e di impatto sulle traiettorie della Strategia Nazionale di Specializzazione Intelligente

a. *Aerosols Clouds and Traces gases Research Infrastructure Network* (di seguito, anche solo ACTRIS - RI);

b. *Central European Research Infrastructure Consortium*, costituito in ERIC con decisione di esecuzione della Commissione europea 2014/392/UE del 24 giugno 2014 (di seguito, anche solo CERIC-ERIC);

c. *Digital Research Infrastructure for the Arts and Humanities*, costituito in ERIC con decisione di esecuzione della Commissione europea 2014/526/UE del 6 agosto 2014 (di seguito, anche solo DARIAH-ERIC);

d. *Distributed High Throughput Computing and Storage - DHTCS* (ora confluito nel progetto IPCEI-HPC-BDA - *Important Project of Common European Interest on High Performance Computing and Big Data enabled Applications*, come documentato nel Piano triennale di attività dell'INFN 2016-2018 e nel FOE 2017 - decreto ministeriale n. 608 dell'8 agosto 2017 - di seguito, anche solo IPCEI-HPC-BDA);

e. *European Carbon Dioxide Capture and Storage Laboratory Infrastructure*, costituito in ERIC con decisione di esecuzione (UE) 2017/996/UE della Commissione del 9 giugno 2017 (di seguito, anche solo ECCSEL-ERIC);

f. *European Life-science Infrastructure for Biological Information* (di seguito, anche solo ELIXIR);

g. *European Marine Biological Resource Centre* (di seguito, anche solo EMBRC);

h. *European Multidisciplinary Seafloor and water column Observatory*, costituito in ERIC con decisione di esecuzione (UE) 2016/1757/UE della Commissione del 29 settembre 2016 (di seguito, anche solo EMSO-ERIC);

i. *European Plate Observing System* (di seguito, anche solo EPOS);

j. *The European Research Infrastructure for Imaging Technologies in Biological and Biomedical Sciences (EuroBioImaging)*, di seguito anche solo EuBI);

k. *Integrated Carbon Observation System*, costituito in ERIC con decisione di esecuzione della Commissione europea 2015/2097/UE del 26 ottobre 2015 (di seguito, anche solo ICOS-ERIC);

l. *European Research Infrastructure for Heritage* (di seguito, anche solo E-RHIS);

m. *Cubic Kilometre Neutrino Telescope* (di seguito, anche solo KM3-NET);

n. *e-Science European Infrastructure for Biodiversity and Ecosystem Research*, costituito in ERIC con decisione di esecuzione della Commissione europea 2017/499/UE del 17 marzo 2017 (di seguito, anche solo LIFEWATCH-ERIC);

o. Laboratori nazionali del Gran Sasso (di seguito, anche solo LNGS);

p. Laboratori nazionali del Sud (di seguito, anche solo LNS);

q. Sardinia Radio Telescope (di seguito, anche solo SRT);

r. *Southern Europe Thomson Back-Scattering Source for Applied Research* (di seguito, anche solo STAR).



PON RICERCA E INNOVAZIONE 2014-2020



FINANZIAMENTI

Dotazione Finanziaria complessiva: 286.094.904,00 € +
40.000.000,00 € (FSC)

- Area in transizione: 34.083.920,00 €
- Area meno sviluppata: 252.010.984,00 €
- Fino al 15% per investimenti localizzati al di fuori dell'Area del programma a condizione che sia garantito che si abbiano ricadute positive nelle Aree del programma in termini di sinergie scientifiche ed investimenti funzionali al rafforzamento del tessuto economico locale.

Finanziamento totale accordato: 293.838.675,16 €



PON RICERCA E INNOVAZIONE 2014-2020



Durata massima dei progetti: 32 mesi - scadenza febbraio 2022
(a meno di deroghe, 4 mesi previsti dal bando + 6 COVID)

- **Spese ammissibili**

- Strumentazione scientifica, apparecchiature e macchinari di ricerca
- Ampliamento di strumentazione scientifica, apparecchiature e macchinari esistenti
- Impianti tecnici funzionali al progetto
- Licenze software o brevetti

- **Spese non ammissibili**

- Servizi
- Personale, però
- nel 2020 bando per reclutare personale. Approvato PON IBISCO CIR 011

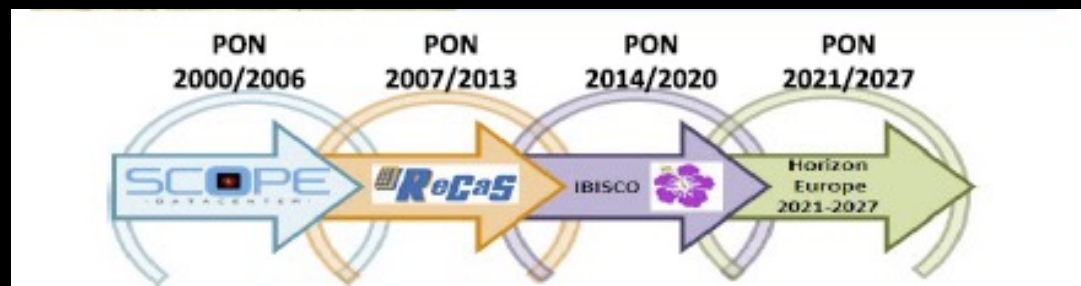


OBIETTIVO DEL PROGETTO IBISCO



Potenziamento dell'infrastruttura RECAS di calcolo scientifico del Sud Italia (INFN BA, CT, NA e UNIBA, UNINA)

- aggiornamento o potenziamento degli impianti dei data centre costituiti con i PON precedenti (Scope e RECAS)



- aumento e aggiornamento delle risorse di calcolo e storage disponibili per mantenere l'infrastruttura competitiva nel contesto internazionale
- Incremento della connettività tra i siti e verso le infrastrutture nazionali e internazionali



OBIETTIVO DEL PROGETTO IBISCO



- **Evoluzione dell'attuale infrastruttura verso una federazione di Data Centre**
 - Facilità per gli utenti di accesso ai dati, senza conoscere la loro collocazione, e ed esecuzione delle applicazioni in maniera trasparente
 - integrazione in modo sinergico e trasparente dei servizi GRID e CLOUD per tutte le esigenze del settore scientifico e non
- **Multidisciplinarietà e multifunzionalità**
 - Andare oltre le attività core INFN aprendosi alla collaborazione con altri ambiti scientifici. La piattaforma IBISCO costituirà il primo esempio di collaborazione attività multi-ente per la condivisione di un'infrastruttura di calcolo multidisciplinare e multifunzionale in grado di adattarsi alle diverse esigenze
 - Il progetto DHTCS, finanziato dal MIUR nel 2013, in collaborazione con CNR, GARR, INAF e INGV ha posto le basi per lo sviluppo di una tale infrastruttura



PARTNER DEL PROGETTO IBISCO PIR 011

Partner RECAS

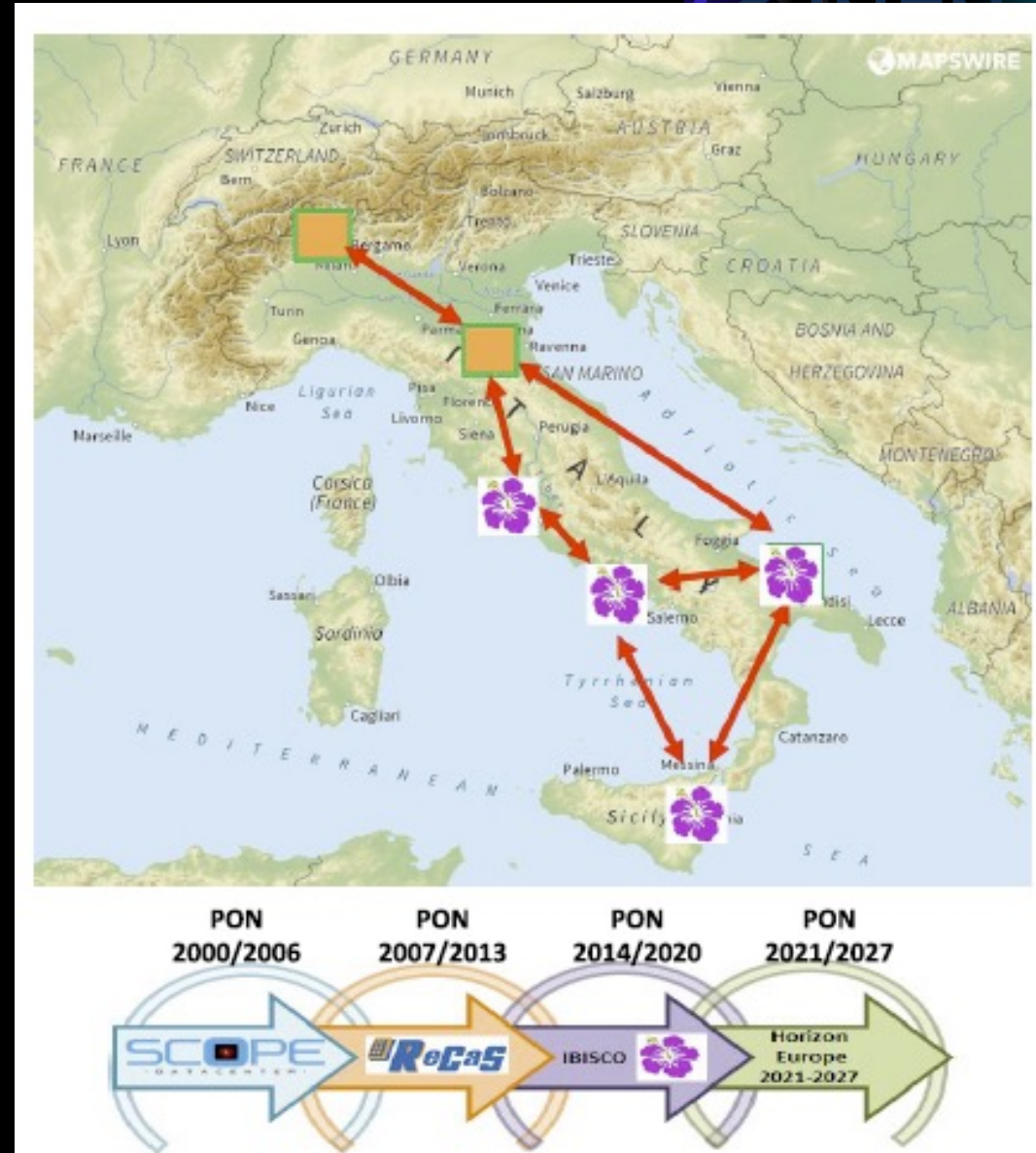
- INFN – Sezioni Bari, Catania, Napoli
- Università – Bari e Napoli

Partner DHTCS

- CNR - Istituti IREA (Bari), ISASI e SPIN (Napoli)
- INAF - Osservatorio di Monteporzio Catone (Laboratori di Frascati)
- INGV - Sezione Roma2 (Laboratori del Sud)

Altre sedi INFN – funzionali allo sviluppo del progetto per la collaborazione con gli altri ER

- Laboratori di Frascati (INAF) e del Sud (INGV)





RIPARTIZIONE FONDI TRA PARTNER E SEDI



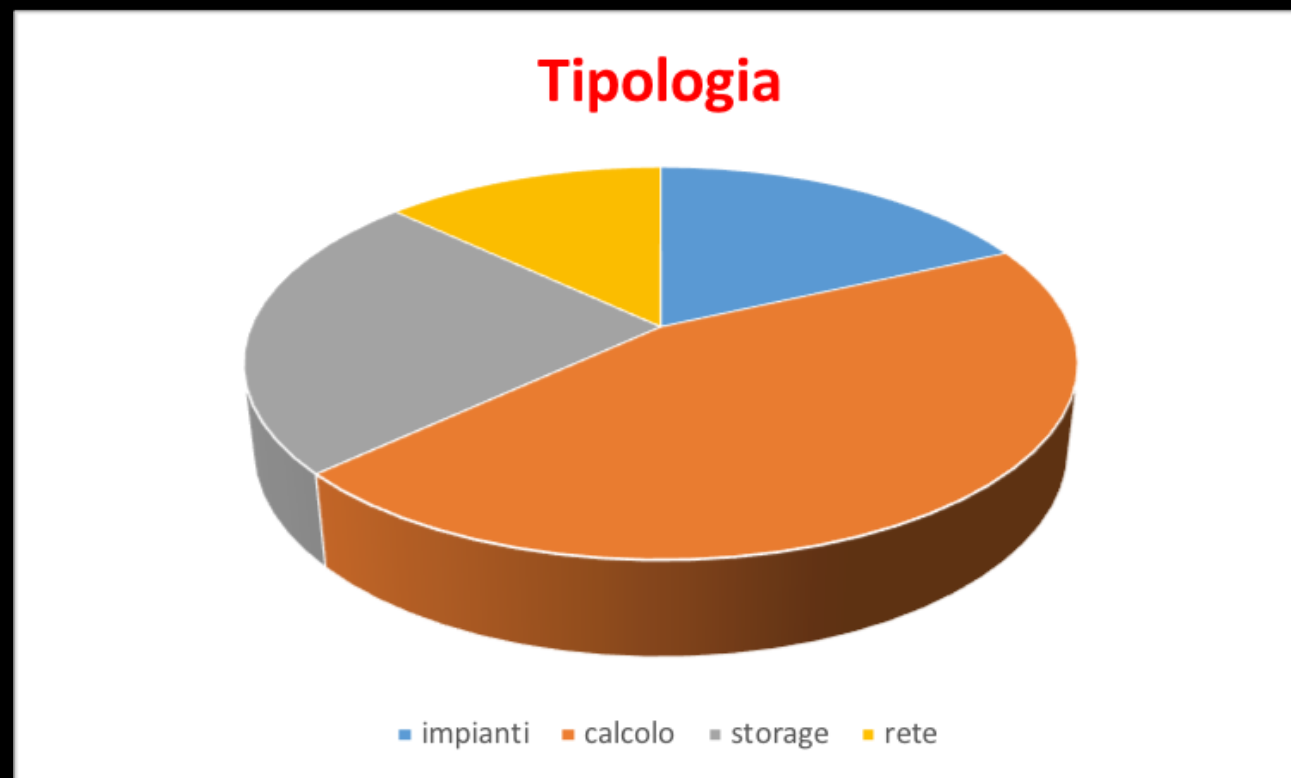
	quote	subtotali		
INFN-BA	23,5%	€	4.394.040,00	€ 11.945.680,00
INFN-CT	17,3%	€	3.234.730,00	
INFN-NA	22,2%	€	4.160.410,00	
INFN-LNF	0,8%	€	156.720,00	
CNR-SPIN	2,8%	€	517.360,00	€ 1.628.210,00
CNR-ISASI	2,8%	€	516.550,00	
CNR-IREA	3,2%	€	594.300,00	
INAF	2,7%	€	502.200,00	€ 502.200,00
INGV	0,8%	€	150.300,00	€ 150.300,00
UNINA	12,3%	€	2.302.670,00	€ 2.302.670,00
UNIBA	11,6%	€	2.170.470,00	€ 2.170.470,00
			€ 18.701.750,00	



RIPARTIZIONE FONDI TRA TIPOLOGIE



	quote	totali	
Impianti	18,1%	€	3.388.890
Calcolo	45,0%	€	8.416.730
Storage	23,9%	€	4.461.800
Rete	13,0%	€	2.434.330





SVILUPPO DEL PROGETTO

Il finanziamento è suddiviso in circa 80 schede di beni da acquistare.

- L'attività principale del progetto consiste nell'acquistare i beni approvati
- → GARE (convenzioni consip, RdO mepa, acquisti diretti, procedure aperte)
- 44 procedure da 20 k€ a 2 M€

UNINA	Gara_UNINA_IMP	impianti
	Conv_UNINA_CPU_1	CPU
	Conv_UNINA_CPU_2	
	Conv_UNINA_Serv	server
	RdO_UNINA_UPS	UPS
	Gara_UNINA_IT	GPU, Storage e Rete
UNIBA	Gara_UNIBA_IMP	Impianti
	Gara_UNIBA_LIB	Libreria
	Gara_UNIBA_CAL	Calcolo
	Gara_UNIBA_NET	Rete
	RdO_UNIBA_IMP	Antincendio
CNR	Gara_CNR_IREA	Calcolo IREA
	Gara_CNR_SP&IS	Calcolo SPIN/ISASI
	RdO_CNR_Cab	Cablaggio ISASI
INAF	Gara_INAF	Calcolo e Storage
INGV	RdO_INGV	Storage

INFN	Gara_INFN	Rete
INFN BA	Gara_BA_IMP	UPS
	Gara_BA_CAL	calcolo
	Gara_BA_STO	storage
	RdO_BA_STO_META	metadata
	RdO_BA_STO_SSD	Storage SSD
	RdO_BA_STO_CEPH	Storage CEPH
	RdO_BA_RACK	rack
	RdO_BA_PDU	pdu
	Diretto_BA_NET	firewall
	RdO_BA_NET	switch
INFN NA	Conv_BA_Serv	server
	Conv_NA_CPU	CPU
	Gara_NA_CAL	calcolo
	Conv_NA_CPU_MC	CPU Many Core
	Conv_NA_Serv	server
	RDO_NA_TC	telecontrollo
	Diretto_NA_RACK	Rack
	Diretto_NA_NET	Switch
	RdO_NA_NET	line card
	Gara_NA_GPU	GPU
INFN CT	Gara_NA_STO	storage
	Gara_CT_IMP	Impianti
	Conv_CT_CPU	CPU
	Gara_CT_IT	Calcolo e Storage
	RdO_CT_VIRT	realtà virtuale
LNF	Conv_CT_NET	switch
	Conv_LNF_Serv	server



OUTPUT DEL PROGETTO



Le Risorse IT - pledge CSNx

Tier2 ufficiali e pledge da MoU

- BA - ALICE, CMS
- CT - ALICE
- NA – ATLAS e Belle2

	2019		2020		2021		Totale		IBISCO	
	CPU	Disco	CPU	Disco	CPU	Disco	CPU	Disco	CPU	Disco
Bari	25.232	1.102	8.440	1.032	24.605	2.438	58.277	4.572	120.000	9.000
Catania	15.987	960	0	242	4.105	160	20.092	1.362	39.000	4.000
Napoli	24.100	910	2.221	0	17.800	896	38.121	1.506	110.000	10.000
Totale annuo	65.319	2.972	10.661	1.274	46.510	3.494	116.490	7.440	269.000	23.000
Ripartizione	75.980	4.246			46.510	3.494				

Considerando il 2022:

- Bari => +10kHS
+2Pbyte
- Napoli
+1Pbyte

NB – è un esercizio: si considerano anche i server HPC e non si assume che tutte le pledge vengano soddisfatte dai siti IBISCO

C'è spazio per altri esperimenti e utenti e attività differenti nei siti INFN



TIPOLOGIA DI RISORSE



- Classici nodi HTC:
 - 128 threads
 - 4GB di Memoria a threads
- Fat Node – HPC – like
 - 256 Core
 - ~2TB of total RAM
 - Up to 3 NVIDIA A100 GPU – card.
 - 6TB SSD local disk



MODALITÀ DI ACCESSO ALLE RISORSE



- HTCondor
 - Fair share based
 - Accesso a singolo host con la possibilità di specificare le risorse necessarie (CPU, RAM, GPU)
- Orchestratore Docker
 - See next slides
- Orchestratore PaaS
 - Approccio INFN-Cloud like
 - Vedi presentazione successiva
- In tutti i casi è necessario registrarsi usando un wizard web molto semplice.

RECAS HPC/GPU CLUSTER: SERVICES

Services ready-to-use:

Interactive remote GPU-based IDE services:

Jupyter Notebook



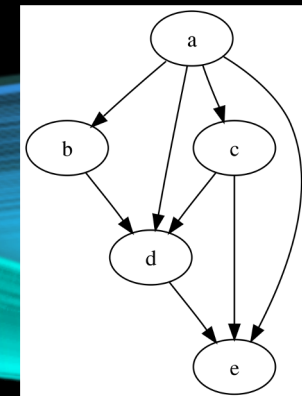
“web service for interactive computing across all programming languages”

Rstudio



“An integrated development environment for R”

- Job Scheduler:
 - Support to GPU-based workflows represented as Directed Acyclic Graphs (DAG)



RECAS HPC/GPU CLUSTER: SERVICES

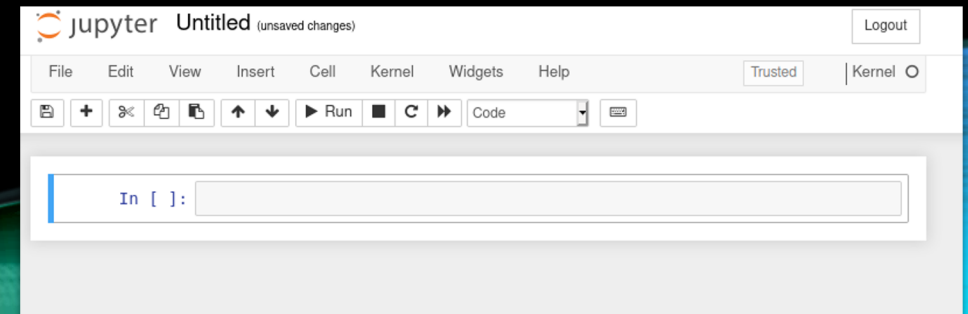
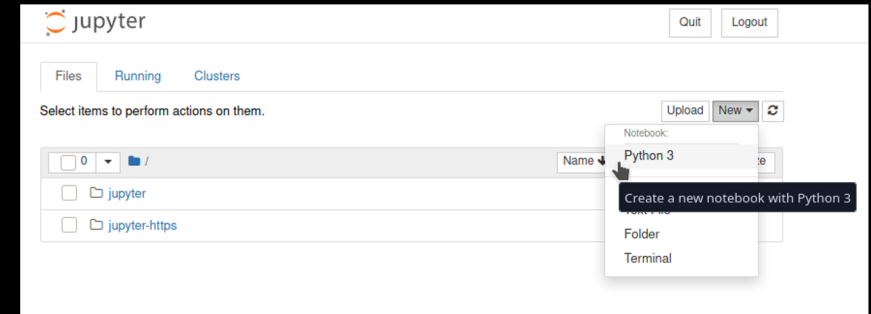
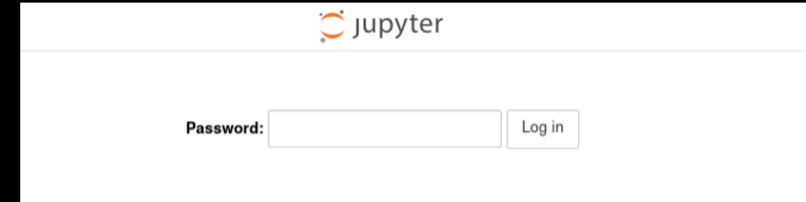
Jupyter Notebook remote IDE

After authentication, users have access to their home directory in the ReCaS distributed storage (GPFS)

Users can immediately create a new Python3 script

The Jupyter IDE (Integrated Development Environment) will be available and users can already write code and execute it

Python modules can be installed directly within the code



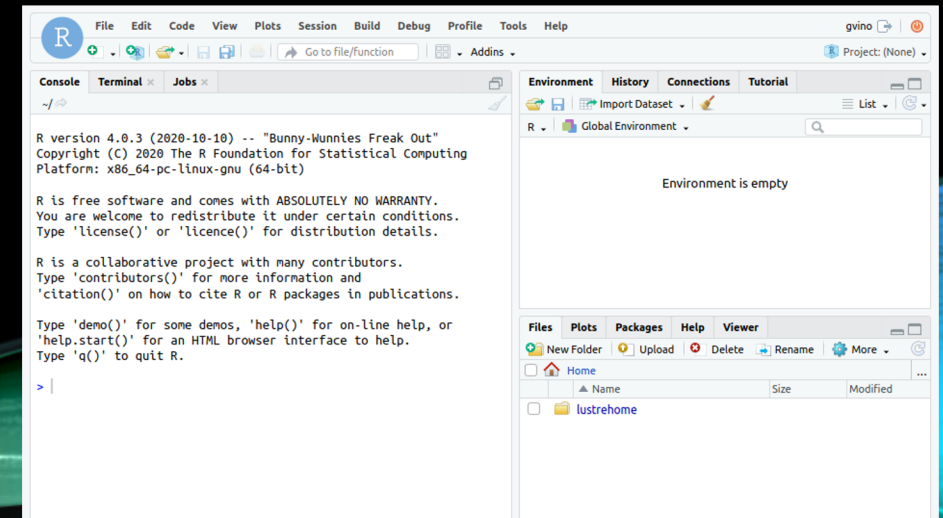
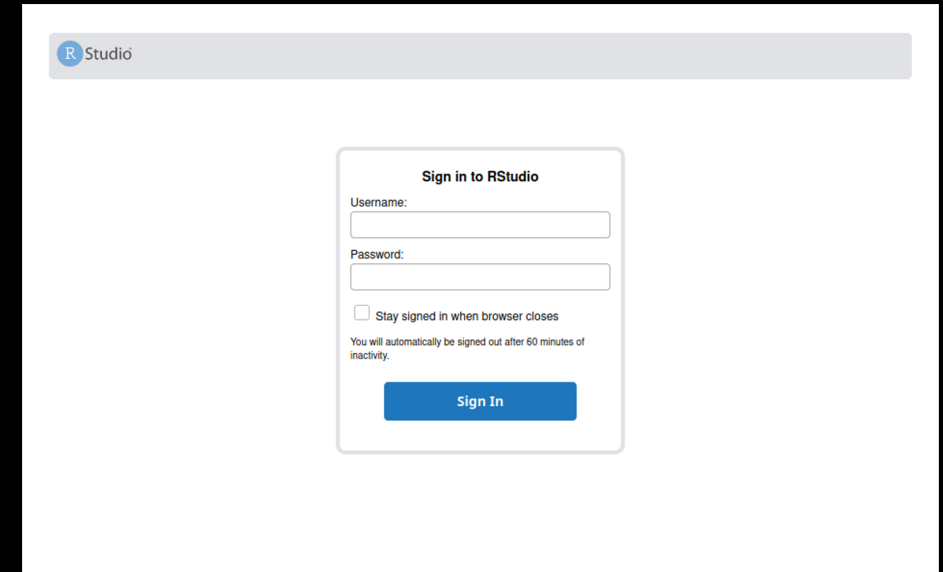
RECAS HPC/GPU CLUSTER: SERVICES

RStudio remote IDE

After authentication, users have access to their home directory in the ReCaS distributed storage (GPFS)

The Rstudio IDE (Integrated Development Environment) will be available and users can already write code and execute it

- R modules can be installed directly within the code



RECAS HPC/GPU CLUSTER: SERVICES

Job Scheduler (Chronos)

Provides an intuitive and simple User Interface (UI) where to check job status

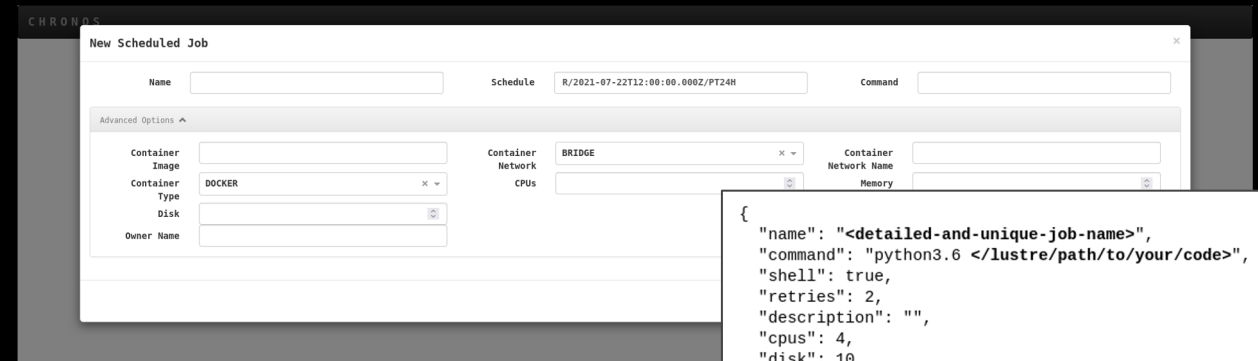
New jobs can be submitted using UI or via command line using a JSON file describing the job

CHRONOS

+ ADD JOB

SUCCESS	0	FAILURE	0	FRESH	1	RUNNING	1	QUEUED	0	IDLE	0
JOB	NEXT RUN		STATUS		STATE		ACTIONS				
gpu-job-gvinoko		in ~365 days		fresh		1 running					

© 2016 Mesosphere, Inc.



CHRONOS

New Scheduled Job

Name Schedule Command

Advanced Options

Container Image Container Type Container Network Container Network Name Container Memory

Owner Name

```
{
  "name": "<detailed-and-unique-job-name>",
  "command": "python3.6 </lustre/path/to/your/code>",
  "shell": true,
  "retries": 2,
  "description": "",
  "cpus": 4,
  "disk": 10,
  "mem": 8192,
  "gpus": 1,
  "environmentVariables": [],
  "arguments": [],
  "runAsUser": "<your-username>",
  "owner": "<your-username>",
  "ownerName": "<your-username>",
  "container": {
    "type": "mesos",
    "image": "<your-container-image>",
    "volumes": [{
      "containerPath":
        "/lustre/path/to/your/home-directory", "hostPath":
        "/lustre/path/to/your/home-directory", "mode": "Rw"
    }]
  },
  "schedule": "R1//PT10S"
}
```

- Manages heterogeneous requests:
 - 2 GPU / 4 CPU / 20 GB RAM
 - 100 CPU / 8GB RAM

RECAS HPC/GPU CLUSTER: UNDER THE HOOD

- **Apache Mesos:**

Unifies all cluster resources in a single virtual entity

Multi-users

High Availability

Manages a lot number of nodes



- **Marathon:**

Runs long running services on top of Apache Mesos

High Availability

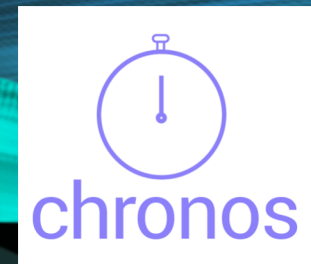
- Load balancing



- **Chronos:**

- Job scheduler for Apache Mesos

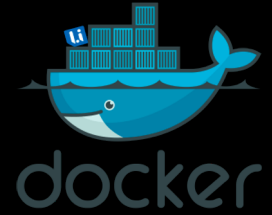
- Supports depending and periodic jobs



RECAS HPC/GPU CLUSTER: UNDER THE HOOD

Docker container:

Contains software, code, libraries and dependencies
Isolates applications from the machine where it is executed
Images are light, standalone and contain all necessary to be run
Official images are available (Nvidia, TensorFlow, ...)



ReCaS HPC/GPU Cluster policy on Docker containers:

- Mandatory for security purpose
- Jupyter Notebook and Rstudio containers have been developed in-house because the majority of the supported use cases needs them
- Not all users' containers can be developed in-house
- An INFN course and a ReCaS tutorial are available to speed-up the user learning process



FUTURO E POSSIBILI SVILUPPI



- Le risorse di ReCaS Bari dovrebbero essere particolarmente adatte agli use case della comunità di AIM.
- È necessario capire come formalizzare l'uso di queste risorse in modo che siano pledged per la comunità
- In altre commissioni è una cosa che viene decisa dai review del calcolo della commissione
- Dovremmo studiare insieme un percorso che permetta di avere questa ufficialità
- Forse il nuovo management del calcolo appena istituito può essere l'opportunità per cominciare questo discorso
- Da un punto di vista tecnologico le cose che sono in sviluppo o in produzione in INFN-Cloud possono essere di interesse per AIM, sia su risorse INFN-Cloud che sia su ReCaS – Bari direttamente