

Cloudnumbers: R en la nube

Carlos J. Gil Bellosta
cgb@datanalytics.com

Datanalytics

Noviembre de 2011

Computación en la nube

Computación en la nube

Consiste en el alquiler de ordenadores por horas

Motivos:

- Acceso a ordenadores más potentes
- HPC, paralelización, etc.

... a un precio asequible.

Cloudnumbers se maneja desde el navegador

Dashboard | My Account | Articles & Support | cloudnumbers.com | Logout
How to get started

Workspaces 1 of 1 used [Create Workspace](#)

ID	Name	Size	Actions
1100	Sample Workspace	16.39 KB	Browse Delete

Remaining included cost: **\$3.3999996**
Upgrade now to unlock the full potential of cloudnumbers.com!
[Invite a friend - 5 hours for free](#)

Sessions [Start old Session](#) [Start Session](#)

ID	Name	Workspace	Start	Application Type	State	Actions
s-e00ee	Prueba Cloudnumbers	Sample Workspace	2011-08-31 01:37:37	R	Installing R 44%	Open Console Refresh Stop Delete

Please note: all terminated sessions will be removed from this grid after 24 hours.

© 2010 - 2011 by cloudnumbers.com. All rights reserved.
OPENFOAM® is a registered trade mark of OpenCFD Limited, the producer of the OpenFOAM software.
RStudio is a trademark of RStudio, Inc.

[cloudnumbers.com](#) | [Legal](#) / [Impressum](#) | [Terms & Conditions](#) | [Contact](#)

Dos áreas: *workspaces* y sesiones

En un *workspace* se almacenan datos que se asocian (montan) en una *sesión*.

Workspaces 1 of 1 used

[Create Workspace](#)

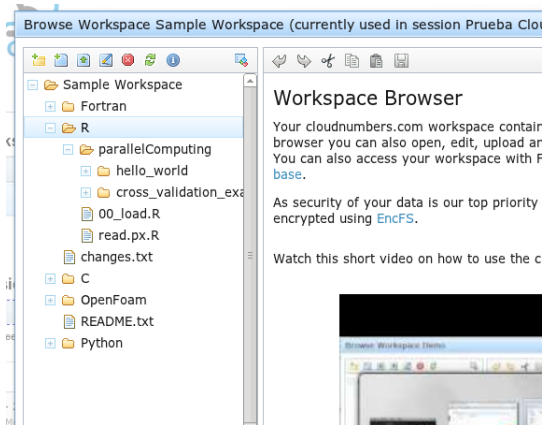
ID	Name	Size	Actions
1100	Sample Workspace	16.39 KB	Browse Delete

Sessions

[Start old Session](#)
[Start Session](#)

ID	Name	Workspace	Start	Application Type	State	Actions
s-e00ee	Prueba Cloudnumbers	Sample Workspace	2011-08-31 01:37:57	R	installing R 44%	Open Console    

Workspaces: subida y bajada de ficheros



Browse Workspace Sample Workspace (currently used in session Prueba Clo

Sample Workspace

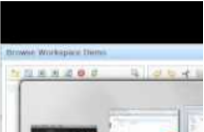
- Fortran
- R
 - parallelComputing
 - hello_world
 - cross_validation_exe
 - 00_load.R
 - read.px.R
 - changes.txt
 - C
 - OpenFoam
 - README.txt
 - Python

Workspace Browser

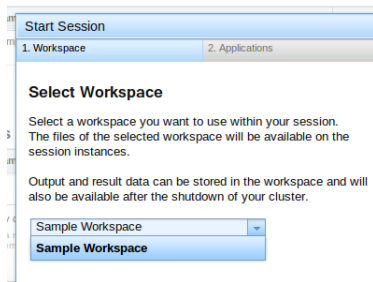
Your cloudnumbers.com workspace contain
browser you can also open, edit, upload ar
You can also access your workspace with F
base.

As security of your data is our top priority
encrypted using [EncFS](#).

Watch this short video on how to use the c



Sesiones 1: selección del *workspace*



Start Session

1. Workspace 2. Applications

Select Workspace

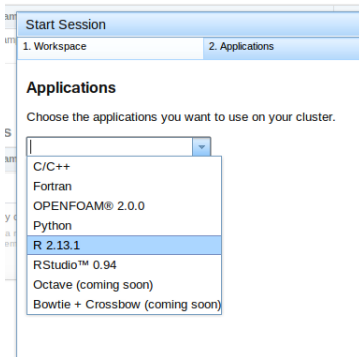
Select a workspace you want to use within your session. The files of the selected workspace will be available on the session instances.

Output and result data can be stored in the workspace and will also be available after the shutdown of your cluster.

Sample Workspace

Sample Workspace

Sesiones 2: selección de aplicaciones



The screenshot shows a 'Start Session' dialog box with two tabs: '1. Workspace' and '2. Applications'. The 'Applications' tab is active. Below the tabs, the text reads 'Applications' and 'Choose the applications you want to use on your cluster.' A dropdown menu is open, showing a list of application options: C/C++, Fortran, OPENFOAM® 2.0.0, Python, R 2.13.1 (highlighted), RStudio™ 0.94, Octave (coming soon), and Bowtie + Crossbow (coming soon).

Start Session

1. Workspace 2. Applications

Applications

Choose the applications you want to use on your cluster.

C/C++

Fortran

OPENFOAM® 2.0.0

Python

R 2.13.1

RStudio™ 0.94

Octave (coming soon)

Bowtie + Crossbow (coming soon)

Sesiones 3: selección de paquetes

The screenshot shows a web interface for starting a session. At the top, there is a navigation bar with four tabs: '1. Workspace', '2. Applications', '3. R Packages', and '4. Instance'. The '3. R Packages' tab is currently selected. Below the navigation bar, the main heading is 'Customize R'. The text explains that users can customize the installation of R by selecting packages from CRAN and BioConductor. It also notes that required system packages will be installed automatically. A link for support is provided: support@cloudnumbers.com. There are two main options for package selection: 'Automatically detect needed packages' (highlighted in a blue box) and 'What does this do?'. Below these, there are two dropdown menus: 'Select CRAN Packages:' and 'Select BioConductor Packages:'. A dark button labeled 'R Packages' is positioned to the right of the BioConductor dropdown.

Start Session

1. Workspace 2. Applications 3. R Packages 4. Instance

Customize R

Customize the installation of R on your cluster by specifying the packages that should be available in the CRAN and BioConductor repository can be selected.

Required system packages that are needed for your packages will automatically be installed required OpenGL libraries will be set up for you.

If you are missing a package, please contact us: support@cloudnumbers.com

Automatically detect needed packages What does this do?

Select CRAN Packages:

Select BioConductor Packages: R Packages

Sesiones 4: selección del *hardware*

Start Session

1. Workspace	2. Applications	3. R Packs
--------------	-----------------	------------

Choose an Instance Type

Choose the hardware you want to compute on.

Instance Type:
Threads: 2 (4 GHz)
RAM: 7.5 GB

Cluster Size:

Price: \$0.50/h

The maximum cluster size is 2.
If you want to start **bigger clusters** and calculate on **larger ins**

Sesiones 5: confirmación

Start Session

1. Workspace 2. Applications 3. R Packages 4. Instance Type 5. Confirm

Confirm

Please confirm your selected options. If you want to start the instances as specified below, click the Finish button.

Workspace: Sample Workspace

Databases:

Application: R

Additional R Packages:

Instance Type: Small (2 Threads, 7.5 GB RAM)

Cluster Size: 1

Additionally you can name this session:

Please note that your cluster will automatically be terminated after 180 minutes.

Sesiones 6: instalación del *software*

Workspaces 1 of 1 used

[Create Workspace](#)

ID	Name	Size	Actions
1100	Sample Workspace	16.39 KB	Browse Delete

Sessions

[Start old Session](#)
[Start Session](#)





ID	Name	Workspace	Start	Application Type	State	Actions
s-e00ee	Prueba Cloudnumbers	Sample Workspace	2011-08-31 01:37:57	R	Installing R <div style="border: 1px solid #ccc; width: 100px; height: 15px; background-color: #add8e6; display: flex; align-items: center; justify-content: center;"> 44% </div>	Open Console    

Pasados unos minutos...

Workspaces 1 of 1 used Create Workspace

ID	Name	Size	Actions
1100	Sample Workspace	16.39 KB	Browse Delete

Sessions Start old Session Start Session

ID	Name	Workspace	Start	Application Type	State	Actions
s-e00ee	Prueba Cloudnumbers	Sample Workspace	2011-08-31 01:37:57	R	RUNNING	Open Console    

... y al arrancar, llega el mensaje comercial

BETA

Upgrade Your Subscription

Free	On-Demand	Advanced	Individual
(current pricing scheme)	choose a payment interval	choose a payment interval	
	<input type="text" value="\$0.00/month"/>	<input type="text" value="\$19.99/month"/>	
Continue Free Trial	Upgrade	Upgrade	Contact Us Now
<ul style="list-style-type: none"> ✓ Traffic included: 2 GB ✓ Storage included: 2 GB ✗ Small instances only ✗ Max cluster size: 2 	<ul style="list-style-type: none"> ✓ Instances with up to 34 GB RAM ✓ 5 workspaces ✓ Workspace size: 10000 MB ✓ Max cluster size: 4 ✓ Traffic included: 20 GB ✓ Storage included: 20 GB 	<ul style="list-style-type: none"> ✓ Instances with up to 68 GB RAM ✓ 15 workspaces ✓ Workspace size: unlimited ✓ Max cluster size: 16 ✓ Traffic included: 30 GB ✓ Storage included: 30 GB 	<ul style="list-style-type: none"> ✓ Individual solution ✓ Start clusters with up to 256 instances ✓ Please contact us to get an individual offer
<p style="margin: 0;">MOST POPULAR</p> <hr style="width: 100%; border: 0; border-top: 1px solid #007bff; margin: 5px 0;"/> <p style="margin: 0; color: #007bff; text-align: center;">View prices per instance</p>			
<p style="margin: 0;">If you want to pay by invoice instead of credit card, please contact us.</p> <p style="margin: 0; color: #dc3545;">* All prices shown are net prices (without VAT). In case you do not provide us with your VAT number, you will be charged the German VAT (19%). For further information please visit our knowledge base.</p>			

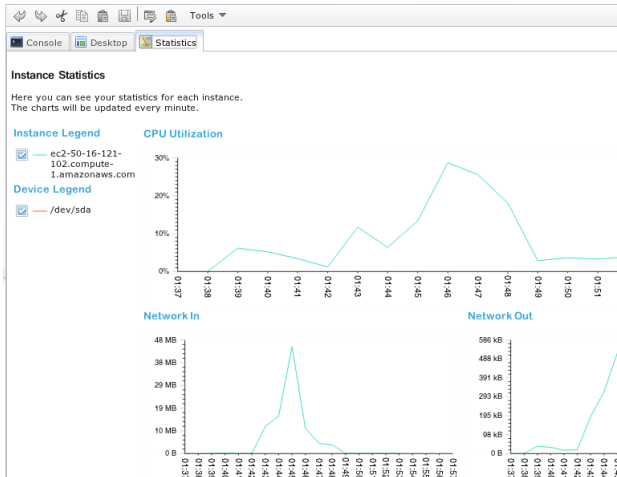
Visión global de la sesión

The screenshot displays the CloudNumbers console interface. At the top, a browser window shows the URL `https://my.cloudnumbers.com/console/s-e00ee`. The main interface is divided into several sections:

- Session Info (Left Panel):**
 - Session State: **RUNNING**
 - Start Date: 2011-08-31 GMT+02:00 01:37
 - Application: R
 - Price: 1x \$0.50/h
 - Buttons: `Show Details`, `Install Packages`
 - Vertical tabs: `Help`, `Upgrade`
 - FAQ section
 - Section: **Frequently asked Questions:**
 - Using an external terminal / SSH console
 - Installing R Packages
 - Using a multicore machine with R
 - Using MPI with R
 - Using MPI or browse the Knowledge Base.
- Desktop Environment (Center):**
 - Windows: `workspace` (active), `ikmarks`, `Help`
 - File Explorer (Right): Shows a tree view of the file system:
 - Sample Workspace
 - Fortran
 - R
 - 00_load.R
 - 01_simulacion.R
 - parallelComputing
 - hello_world
 - cross_validation_example
 - read.px.R
 - changes.txt
 - C
 - OpenFoam
 - README.txt
 - Python

- Bottom Panel:**
- Terminal window: `rstudio-0.94.10.....deb`
- System tray: `Show all downloads...`, `01:57`

Pestaña de estadísticas



Edición de ficheros del *workspace* en su pesaña

The screenshot displays the RStudio environment. The main window shows a script editor with the following R code:

```
#####
# Carga de datos y funciones auxiliares
# cjgb, 201108
#####

#library( pxR )
#library( xtable )

province.numbers <- sapply( 1:52, function( x ) sprintf( "%05d", x ) )
province.numbers <- paste( "http://www.ine.es/pcaxisdl//t20/e245/p05/a2010/10/", province.numbers,
res <- sapply( province.numbers, function( x ) as.data.frame( read.px( url( x ) ) ), simplify = F

total <- do.call( rbind, res )

rm( res, province.numbers ); gc()

colnames( total ) <- c( "edad", "municipio", "sexo", "n" )
rownames( total ) <- NULL

total <- subset( total, n > 0 ) # borra celdas vacías

total <- subset( total, edad != "Total" )
total$edad <- factor( total$edad )

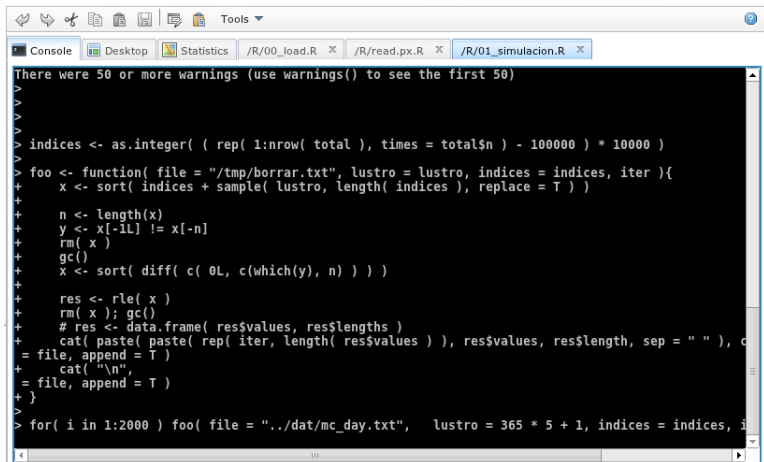
total <- subset( total, sexo != "Ambos sexos" )
total$sexo <- factor( total$sexo )

total <- subset( total, municipio != "Total" )
total$municipio <- factor( total$municipio )
```

The right-hand pane shows the File Explorer with the following structure:

- Sample Workspace
 - Fortran
 - R
 - 00_load.R
 - 01_simulacion.R
 - parallelComputing
 - hello_world
 - cross_validat
 - read.px.R
 - changes.txt
 - C
 - OpenFoam
 - README.txt
 - Python

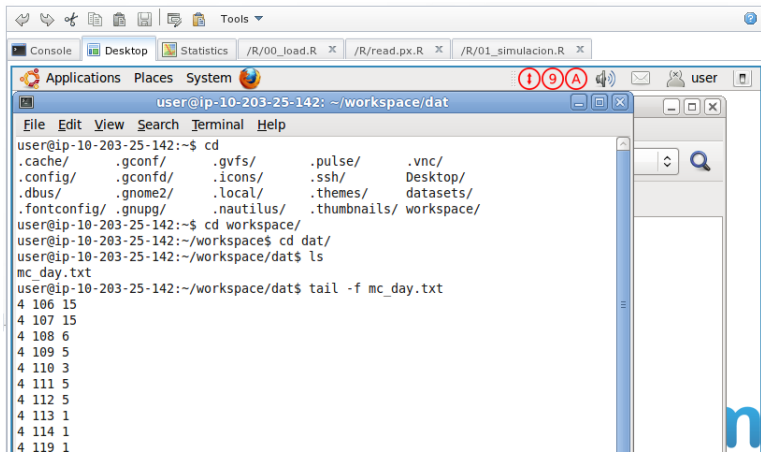
Una de las pestañas es la consola...



The screenshot shows an R console window with several tabs: Console, Desktop, Statistics, /R/00_load.R, /R/read.px.R, and /R/01_simulacion.R. The console output is as follows:

```
There were 50 or more warnings (use warnings() to see the first 50)
>
>
>
> indices <- as.integer( ( rep( 1:nrow( total ), times = total$n ) - 100000 ) * 10000 )
>
> foo <- function( file = "/tmp/borrar.txt", lustro = lustro, indices = indices, iter ){
+   x <- sort( indices + sample( lustro, length( indices ), replace = T ) )
+
+   n <- length(x)
+   y <- x[1:n] != x[-n]
+   rm( x )
+   gc()
+   x <- sort( diff( c( 0L, c(which(y), n) ) ) ) )
+
+   res <- rle( x )
+   rm( x ); gc()
+   # res <- data.frame( res$values, res$lengths )
+   cat( paste( paste( rep( iter, length( res$values ) ), res$values, res$length, sep = " " ), c
= file, append = T )
+   cat( "\n",
= file, append = T )
+ }
>
> for( i in 1:2000 ) foo( file = "../dat/mc_day.txt", lustro = 365 * 5 + 1, indices = indices, i
```

... pero se puede abrir otra de Ubuntu así...



The screenshot shows a desktop environment with a terminal window open. The terminal window title is "user@ip-10-203-25-142: ~/workspace/dat". The terminal output shows the following commands and results:

```
user@ip-10-203-25-142:~$ cd
.cache/      .gconf/     .gvfs/      .pulse/     .vnc/
.config/     .gconfd/   .icons/     .ssh/       Desktop/
.dbus/       .gnome2/   .local/     .themes/    datasets/
.fontconfig/ .gnupg/    .nautilus/  .thumbnails/ workspace/
user@ip-10-203-25-142:~$ cd workspace/
user@ip-10-203-25-142:~/workspace$ cd dat/
user@ip-10-203-25-142:~/workspace/dat$ ls
mc_day.txt
user@ip-10-203-25-142:~/workspace/dat$ tail -f mc_day.txt
4 106 15
4 107 15
4 108 6
4 109 5
4 110 3
4 111 5
4 112 5
4 113 1
4 114 1
4 119 1
```

... o así

The screenshot shows a remote desktop window titled "user@ip-10-203-25-142: ~". Inside the window, a terminal application is open, displaying the output of the 'top' command. The terminal output shows system statistics and a list of processes.

System Statistics:

```
top - 00:21:12 up 40 min, 3 users, load average: 1.03, 1.14, 0.81
Tasks: 148 total, 2 running, 145 sleeping, 0 stopped, 1 zombie
Cpu(s): 13.2%us, 19.4%sy, 0.0%ni, 65.8%id, 0.0%wa, 0.0%hi, 0.0%si, 1.7%st
Mem: 7645964k total, 2075748k used, 5570216k free, 121120k buffers
Swap: 0k total, 0k used, 0k free, 715892k cached
```

Process List:

PID	USER	PR	NI	VIRT	RES	SHR	S	%CPU	%MEM	TIME+	COMMAND
9565	user	20	0	984m	750m	3572	R	100	10.0	14:58.84	R
937	root	20	0	110m	38m	2896	S	0	0.5	0:19.07	ruby
13511	user	20	0	231m	14m	10m	S	0	0.2	0:00.31	gnome-terminal
1	root	20	0	24124	2284	1336	S	0	0.0	0:00.43	init
2	root	20	0	0	0	0	S	0	0.0	0:00.00	kthreadd
3	root	20	0	0	0	0	S	0	0.0	0:01.73	ksoftirqd/0
4	root	20	0	0	0	0	S	0	0.0	0:00.00	kworker/0:0
5	root	20	0	0	0	0	S	0	0.0	0:00.01	kworker/u:0
6	root	RT	0	0	0	0	S	0	0.0	0:00.00	migration/0
7	root	RT	0	0	0	0	S	0	0.0	0:00.00	migration/1
8	root	20	0	0	0	0	S	0	0.0	0:00.00	kworker/1:0
9	root	20	0	0	0	0	S	0	0.0	0:00.56	ksoftirqd/1
10	root	0	-20	0	0	0	S	0	0.0	0:00.00	cpuset
11	root	0	-20	0	0	0	S	0	0.0	0:00.00	khelper
12	root	0	-20	0	0	0	S	0	0.0	0:00.00	netns

Y al terminar...

... uno cierra su sesión...

Stop Session

... y puede descargar los ficheros de resultados de su *workspace*.

¡Fin!