

Istruzioni generali per l'utilizzo delle macchine google

La creazione e modifica di macchine virtuali si gestisce attraverso la GCP console, all'indirizzo <https://console.cloud.google.com/compute/> e nella sezione "MV-instances"

Creazione chiave ssh

Per creare una chiave ssh, utilizzare il comando

```
ssh-keygen -t rsa -C "studente"
```

Quando viene richiesto il passphrase, date invio per lasciare vuoto. Si creano due file: una chiave privata, `id_rsa` ed una chiave pubblica, `id_rsa.pub`. La chiave pubblica va caricata sulla macchina virtuale. Per fare questo:

- clic sul nome della macchina (e.g. `instance-1`)
- clic su 'MODIFICA' o (EDIT)
- scorrere fino alla sezione SSH Keys.
- incollare la chiave pubblica, ossia `id_rsa.pub`.
- se utilizzate la macchina creata con il disco visto a lezione, controllare che la chiave pubblica incollata termini con 'studente'

Login

```
ssh studente@ipaddr
```

dove `ipaddr` rappresenta l'indirizzo ip associato alla macchina (External ip). Ad esempio,

```
ssh studente@35.121.33.11
```

Di default `ssh` utilizza le chiavi contenute in `~/.ssh`. In alcuni casi, potrebbero essere presenti chiavi multiple. In tal caso, bisogna specificare in modo opportuno la chiave privata associata alla chiave pubblicata caricata sulla macchina,

```
ssh studente@35.121.33.11 -i ~/.ssh/id_rsa
```

sostituendo ad `~/.ssh/id_rsa` il percorso opportuno.

Una volta effettuato il login, il nome utente sulla shell cambia. Ad esempio, avrete qualcosa del tipo

```
[studente@instance-1]:~$
```

che suggerisce la buona riuscita del login.

Rstudio server

Una volta lanciata l'istanza, aprire il browser all'indirizzo `http://ipaddr:80` (nota che l'indirizzo inizia con `http`, NON `https`, e finisce per `: 80`). Ad esempio, `http://35.121.33.11:80`.

Rstudio server dovrebbe partire di default come servizio all'accensione. In caso contrario, fate il login con `ssh` ed eseguite i seguenti comandi

```
sudo rstudio-server start  
sudo rstudio-server status
```

Le credenziali di accesso sono

```
user: studente  
password: password
```

In caso di problemi di login, è sufficiente cambiare la password associata all'account studente. Per fare questo, da dentro la macchina,

```
sudo passwd studente
```

e digitate due volte la nuova password (lo schermo non mostra le lettere che inserite). In caso di password troppo semplici (come appunto `password`), il sistema ci mostra un errore, che tuttavia è possibile ignorare senza problemi.

Pyspark

Per avviare pyspark con le impostazioni di default, una volta effettuato il login lanciare semplicemente

```
pyspark
```

Si aprirà un ambiente python 3.5 interattivo e collegato a spark.

In alcuni casi, spark potrebbe richiedere la configurazione manuale di alcuni parametri per funzionare al meglio. Per una lista completa delle opzioni, si veda <https://spark.apache.org/docs/latest/configuration.html>.

Ad esempio, per eseguire lo script che trovate su moodle, il comando da lanciare è

```
pyspark --driver-memory="10G"
```

Nella maggior parte dei casi, il comando sopra dovrebbe essere sufficiente ad avere risorse sufficienti per analizzare il vostro dataset.

sparkRd

Per avviare sparkR, lanciare R o Rstudio normalmente ed utilizzare i comandi nello script `sparkR.R` su moodle. Le opzioni importanti sono le seguenti

```
Sys.setenv(SPARK_HOME = "/home/spark")  
library(SparkR, lib.loc = c(file.path(Sys.getenv("SPARK_HOME"), "R", "lib")))  
sparkR.session(master = "", sparkConfig = list(spark.driver.memory = "10g"))  
sparkR.conf()
```

Le opzioni configurabili all'interno di `sparkR.session` sono le medesime citate sopra con `pyspark`. La libreria `sparkR` contiene tutte le funzioni necessarie ad interfacciare R e spark.