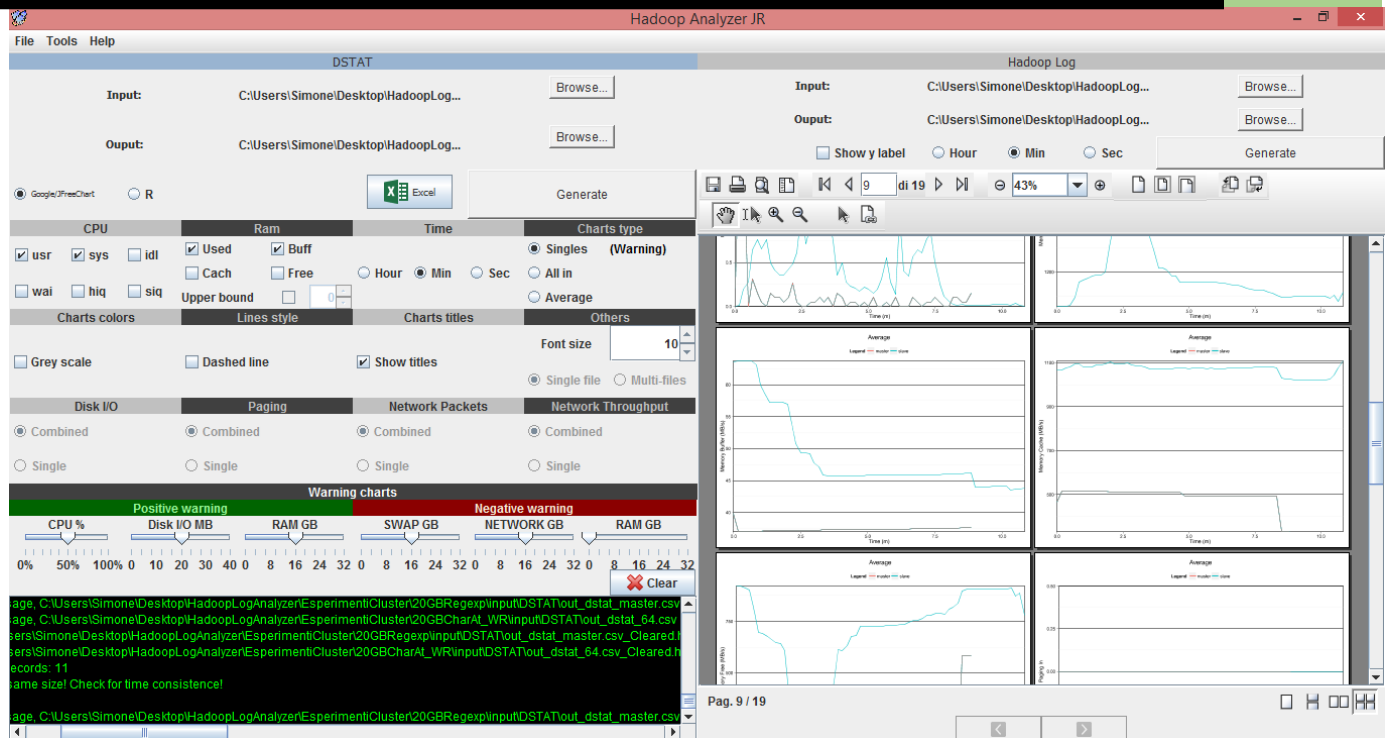


2014/2015

Hadoop Analyzer JR User manual



Amedeo Leo – Alessio Petrozziello –
Simone Romano
Progetto Sistemi Operativi Avanzati
2014/2015

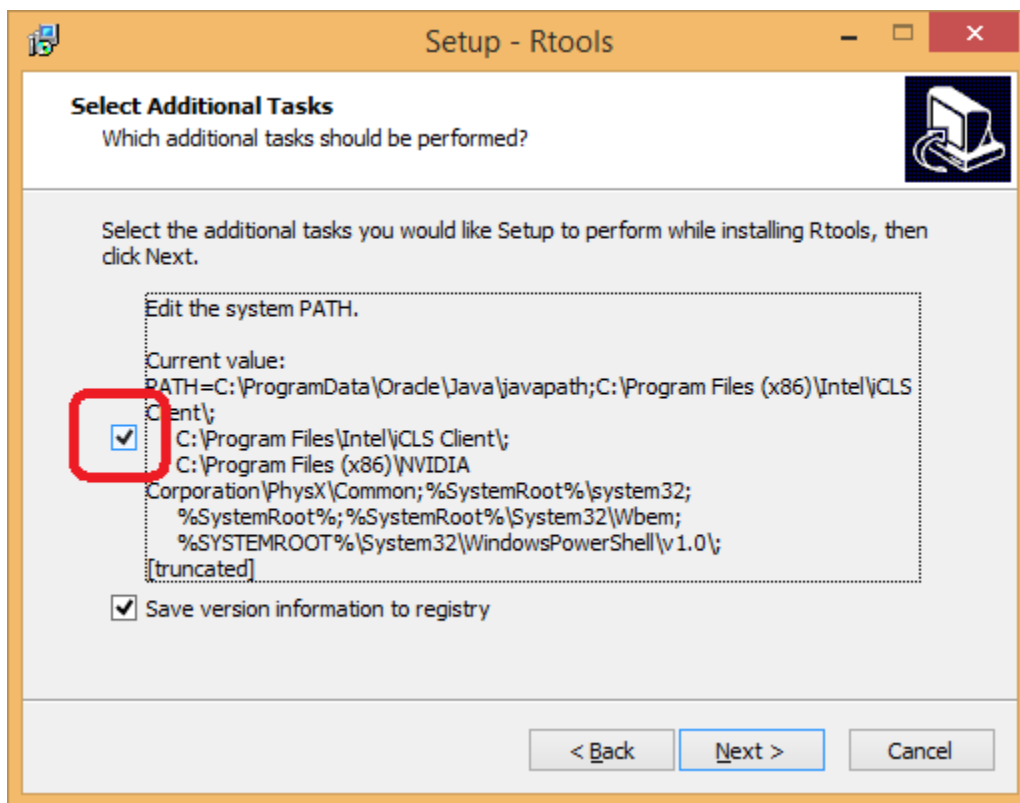
Installazione

Per iniziare, installare R dal sito <http://cran.r-project.org> (scegliere la versione binaria base).

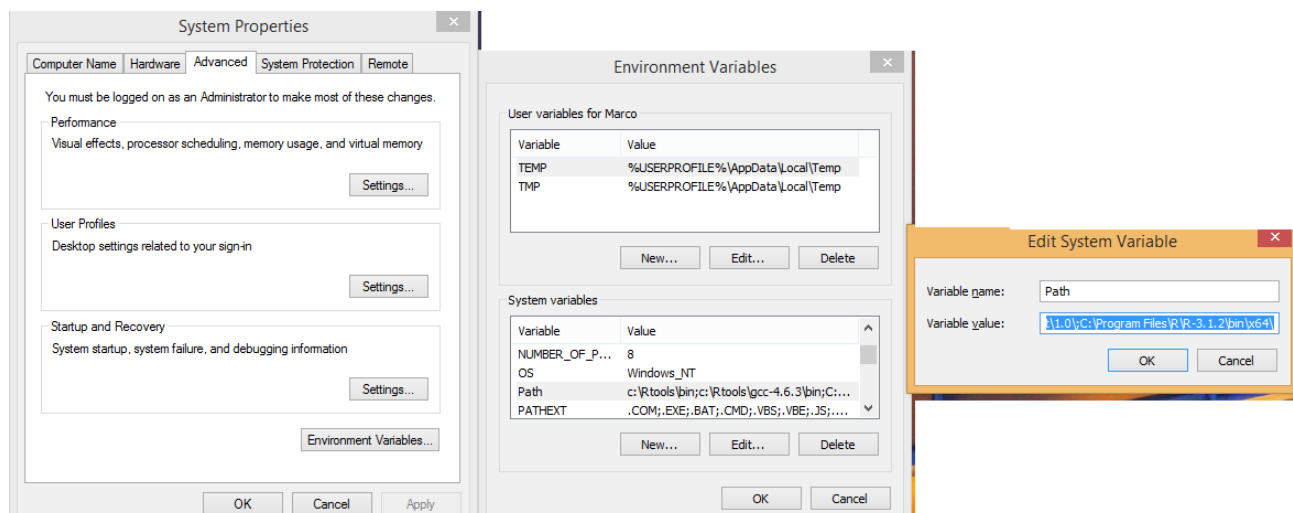
Per utenti **Windows**:

Installare Rtools dallo stesso sito.

Occorre fare attenzione durante la fase di installazione di Rtools. È necessario procedere come di seguito:



Terminata l'installazione, settare la variabile PATH facendola "puntare" alla cartella dei file di R. Ad esempio,



Installato Rtools, lanciare il file 'install.bat' presente nella cartella 'HadoopLogAnalyzer'.

Per utenti **Mac**:

lanciare il file 'install.sh' presente nella medesima cartella. Eventualmente assicurarsi che tale file abbia i permessi di esecuzione.

E' possibile settare tali permessi con:

```
chmod +x install.sh
```

Error

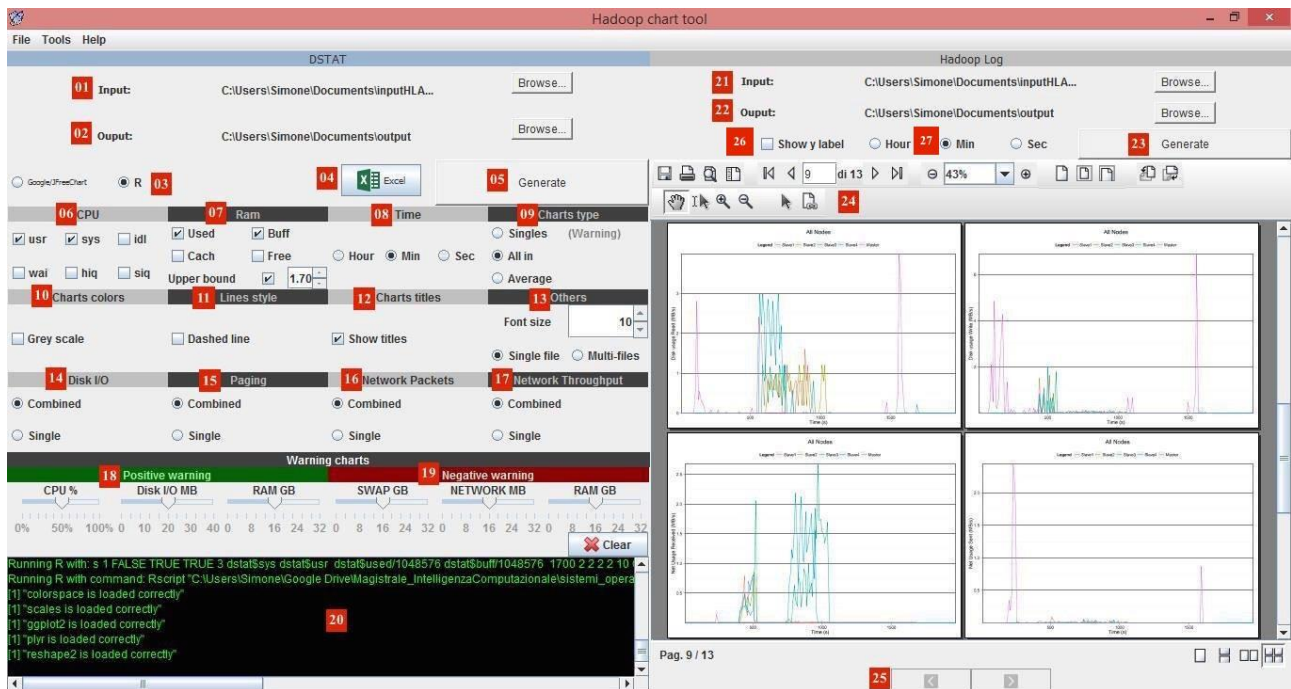
1. Nel caso in cui 'R' non fosse installato sulla macchina verrà mostrato a video il seguente errore:

```
Saving file: C:\Users\Marco\Desktop\myConf.hlaConf
Start...
Running R with: s 1 FALSE TRUE TRUE 1 NULL NULL NULL 2 2 2 10 C:\Users\Marco\Desktop\HadoopLogAnalyzer\inputH
Running R with command: Rscript "C:\Users\Marco\Desktop\HadoopLogAnalyzer\R_script\chart_generator.R"
Exception executing r_script: Cannot run program "Rscript": CreateProcess error=2, The system cannot find the file specified
```

Si può risolvere installando R come spiegato precedentemente.

Manuale utente

Il tool si presenta con questa interfaccia.



Come si può notare, è divisa in dstat, sulla sinistra, e Hadoop log, sulla destra. Di seguito sono descritte le diverse caratteristiche.

01: Inserire la cartella di input dei file di dstat più il file “prof_output.txt” eventualmente generato con l’attivazione del profiler. Per abilitare il profiler bisogna passare alla JVM la seguente riga:

```
-agentlib:hprof=cpu=times,heap=sites,depth=6,interval=10000,force=n,thread=y,verbose=n,file=prof.output
```

02: Inserire la cartella di output.

03: Scegliere il tool da utilizzare per generare i grafici (“R” creerà dei pdf, “Google/JFreeChart” una pagina web).

04: Creazione del file excel contenente l’aggregazione dei dati generati da dstat.

05: Generazione dei grafici di dstat.

06: Possibilità di scegliere tra le varie percentuali di utilizzo della CPU. Nel caso in cui “Google/JFreeChart” sia spuntato, saranno mostrati un grafico per ciascuna delle sei scelte e un altro grafico per la somma delle percentuali scelte dall’utente. Nel caso in cui “R” sia spuntato, se nessuna scelta è selezionata, saranno generati dei grafici per ciascuna scelta; se due o più scelte sono selezionate, sarà mostrato all’utente un grafico contenente la somma delle scelte effettuate.

07: Possibilità di scegliere tra le varie combinazioni della RAM. Nel caso in cui “Google/JFreeChart” sia spuntato, saranno mostrati un grafico per ciascuna delle quattro scelte e un altro grafico per la somma delle scelte dell’utente. Nel caso in cui “R” sia spuntato, se nessuna scelta è selezionata, saranno generati dei grafici per ciascuna scelta; se due o più scelte sono selezionate, sarà mostrato all’utente un grafico contenente la somma delle scelte effettuate; inoltre viene data la possibilità di inserire un upperbound al grafico della RAM.

08: Scelta della timeline in ore, minuti o secondi.

09: Scelta del tipo di grafico: nel caso in cui “Google/JFreeChart” sia spuntato, saranno mostrati i grafici in una pagina web, in cui sono mostrati anche i grafici di warning. Nel caso in cui “R” sia spuntato, saranno mostrati i grafici in dei file pdf. La scelta è tra “Singles”, “All in” e “Average”. Nel primo caso, se “Google/JFreeChart” è spuntato, verranno prodotti dei grafici singoli in cui sarà mostrato il relativo slave e il master per ciascuna delle scelte effettuate; se invece “R” è spuntato, saranno generati dei grafici singoli per ciascuno slave e per il master a seconda delle scelte fatte. Nel secondo caso, viene generato un solo grafico comprensivo di tutti i nodi. Nel terzo, viene fatta la media per ciascuna delle scelte effettuate.

10: Scelta della colorazione del grafico (scala di grigi, di default è a colori).

11: Scelta dello stile della linea (tratteggiata, di default è continua).

12: Possibilità di inserire il titolo in ciascun grafico.

13: Altre possibilità date all’utente: dimensione del font e scelta del tipo di grafico, su un singolo file o su più file; nel secondo caso, sarà possibile effettuare la scelta solo se “R” è stato spuntato.

14: Possibilità di scegliere se combinare in uno stesso grafico l’I/O del disco o in due grafici separati. Sarà possibile effettuare la scelta solo se “R” è stato spuntato.

15: Possibilità di scegliere se combinare in uno stesso grafico il paging o in due grafici separati. Sarà possibile effettuare la scelta solo se “R” è stato spuntato.

16: Possibilità di scegliere se combinare in uno stesso grafico i pacchetti ricevuti o inviati o in due grafici separati. Sarà possibile effettuare la scelta solo se “R” è stato spuntato.

17: Possibilità di scegliere se combinare in uno stesso grafico l’utilizzo della rete in e out o in due grafici separati. Sarà possibile effettuare la scelta solo se “R” è stato spuntato.

18: Nel caso in cui “Google/JFreeChart” sia attivato, l’utente può settare, per i grafici di warning positivi, la percentuale di CPU, di utilizzo del disco (in MB/s) e della RAM (in GB). Se l’utilizzo di CPU è minore della percentuale inserita dall’utente, se il disco è stato utilizzato meno di quanto ha inserito l’utente e se la RAM è stata utilizzata meno di quanto inserito dall’utente, allora, per ciascuno slave in cui tali assunzioni risultano vere, verrà segnalato un warning positivo.

19: Nel caso in cui “Google/JFreeChart” sia attivato, l’utente può settare, per i grafici di warning negativi, lo swap (in MB), l’utilizzo della rete (in MB/s) e della RAM (in GB). Se almeno una delle risorse è utilizzata più di quanto l’utente ha selezionato allora verrà dato un warning negativo.

20: Shell che mostra la computazione del tool.

21: Inserire la cartella di input dei log di Hadoop.

22: Inserire la cartella di output dei log di Hadoop.

23: Generazione dei grafici dei log di Hadoop.

In caso di errore controllare che nella cartella di input dei log non siano presenti file diversi da ‘attempt*’.

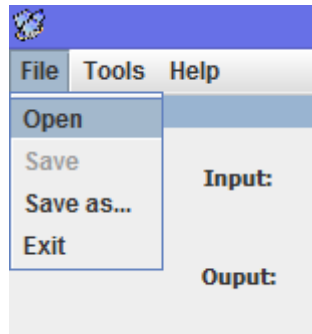
24: Pdf viewer per i grafici generati dal linguaggio R.

25: Possibilità di visualizzare il file precedente o successivo all'interno della stessa cartella.

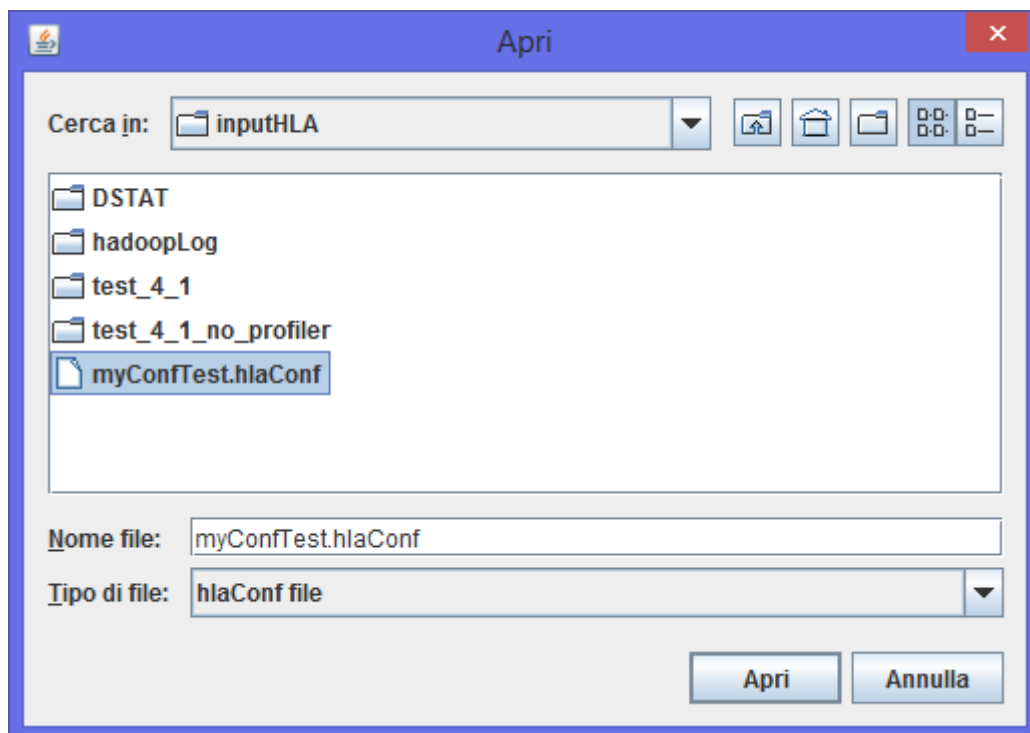
26: Permette di mostrare le label con i nomi degli slave.

27: Scelta della timeline in ore, minuti o secondi.

Nel menù, cliccando su "File", sono presentate queste scelte.

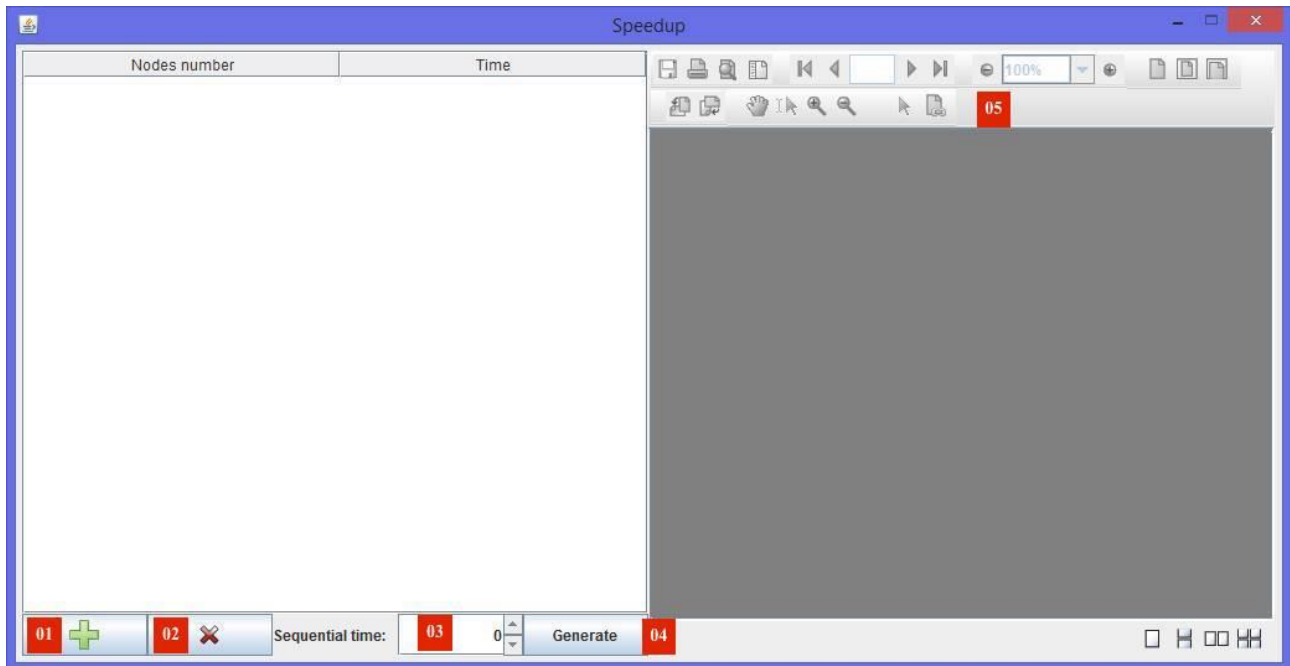


Il tool permette di inserire una configurazione precedentemente salvata dall'utente. Cliccando su "Open", l'utente può scegliere il file precedentemente creato.



Se l'utente desidera salvare la nuova configurazione, cliccando su "Save" sarà automaticamente salvata al posto della precedente. Altrimenti, l'utente deve necessariamente cliccare "Save as..." per poter salvare la configurazione. "Exit" permette di chiudere il programma.

Nel menù, cliccando su “Tools”, si apre questa schermata.



Si permette all’utente di inserire i parametri di seguito descritti, in modo da generare i grafici relativi allo speedup e all’efficienza.

01: Permette di aggiungere una riga alla tabella. In particolare, l’utente può inserire il numero di nodi e il relativo tempo di esecuzione.

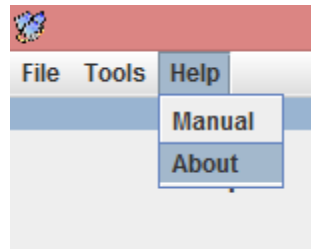
02: Permette di cancellare l’ultima riga della tabella.

03: Permette di inserire il tempo sequenziale.

04: Genera i grafici relativi a speedup e efficienza.

05: Pdf viewer per i grafici generati dal linguaggio R.

Nel menù, cliccando su “Help”, sono visualizzate due opzioni.



Cliccando su “Manual”, sarà scaricato il manuale che si sta leggendo.

Cliccando su “About”, viene visualizzata questa schermata.

