D4Science Infrastructure - Task #10758

R Packages Lists for RStudio - Distinguish between Prototyping and Production Environments

Dec 19, 2017 04:07 PM - Paolo Scarponi

Status:	New		**Start date:**	Dec 19, 2017
Priority:	Normal		**Due date:**	
Assignee:	Roberto Cirillo		**% Done:**	0%
Category:			**Estimated time:**	0.00 hour
Target version:	Algorithms Development Tools Improvements			
Infrastructure:	Production			

Description

We currently have to sets of R packages lists, one for the Prototyping Environments (the VREs with SAI) and one for the Production Environments (the VREs without SAI). The RStudio machines retrieve the packages lists from the production set, but this may cause some inconsistencies when a user is writing his code using RStudio in a Prototyping VRE and then publishes it onto DataMiner using SAI. In fact, he might be able to run his script inside RStudio without errors, while encountering issues once the algorithm is in DataMiner, because one or more of the packages he uses are not actually installed on the machines. In a perfect world the user wouldn't forget to include all the dependencies in SAI before the publication step so to have them automatically installed on the machines, but we do not live in a perfect world and if this happens we automatically have the aforementioned inconsistencies.

This ticket is meant to highlight such issue and find a solution together.

One possible way to tackle the problem consists in a distinction between RStudio machines, having the ones behind Prototyping Environments point to the Prototyping packages list, while the ones behind Production Environments points to the other set. Is it a viable escamotage? Do you have any other ideas?

History

#1 - Dec 20, 2017 01:50 AM - Pasquale Pagano

- Priority changed from Low to Normal

#2 - Dec 20, 2017 11:03 AM - Roberto Cirillo

I think it is the right way for avoiding this kind of issue. The proto instances could be named with the "-proto" suffix in order to better understand if an instance is synchronized with the proto repository or the production repository. @andrea.dellamico@isti.cnr.it what do you think?

#3 - Dec 20, 2017 12:08 PM - Andrea Dell'Amico

It's doable, the R packages management is the same on both the dataminer and rstudio servers.
Renaming a node is a pain but we can add CNAMES. We need to know what VMs run as proto instances and create a group under the ansible playbook with the different configuration.

#4 - Dec 21, 2017 02:24 PM - Roberto Cirillo

It is not that simple, really. There are several rstudio instances that are shared with both proto and production VREs. We should first completely separate the two set of instances and after we can create a separated group under the ansible playbook. In order to separate the two set, I guess we should first transfer the user's home to the new production instances.

#5 - Dec 21, 2017 03:12 PM - Andrea Dell'Amico

I see. If we can wait, there's a chance that the GARR cloud people will enable CephFS for us in some weeks. With it we could mount the same home directory on every rstudio server that runs in the GARR cloud and then no more need to move home directories around.