R on Amazon EC2

Karim Chine[1]

1.Cloud Era Ltd, Cambridge, UK

* Contact author: karim.chine@m4x.org

Keywords: biocep, cloud computing, EC2, distributed computing, R workbench, collaborative data analysis

Biocep[1] builds on top of R an open platform for computing and data analysis. The Amazon Elastic Compute Cloud (Amazon EC2) web service provides users with the ability to execute applications in Amazon's computing environment. Using a rich workbench within the browser, the statistician can now work with an R server running on EC2 as if it was local to his machine. The platform hides the complexity of Amazon's cloud computing infrastructure and the R server is abstracted with a simple URL. multiple statisticians can connect simultaneously to the same EC2-R server and analyze data collaboratively via a set of broadcasted views. For example, the console log is sent in real time to all users. Chatting is enabled and a graphic device is synchronously updated for all. Biocep includes an editable R-enabled collaborative spreadsheet that retains data on the server, removing limits on client machines. Distributed and linked statistical graphics based on a refactored iplots[2] package enable the collaborative highlighting and color brushing of various linked plots.

Biocep makes distributed computing using R and EC2 accessible to a larger number of statisticians. Easy-to-use functions enable the control from within an R session of several EC2-R workers individually or as a cluster to solve embarrassingly parallel problems. The SOAP-R and RESTful-R Biocep's frontends enable the use of pools of EC2-R workers from Perl,Python, C, C#.. A provided web application uses EC2-R pools to expose an API similar to the Google charts API that returns R Graphics in any format in response to a URL.

Once connected to an R server running at any location, the Biocep's workbench enables data analysis applications wrapped as plugins to access to that server. The workbench can be used as a RESTful Web Service bridge between the R Server and various desktop applications (Excel, OpenOffice,..).

The presentation will include demos of some of the described use cases using publicly available Biocep-R Amazon Machine Instances.

References

[1] Karim Chine(2008), "Biocep, Towards a Federative, Collaborative, User-Centric, Grid-Enabled and Cloud- Ready Computational Open Platform,"
escience,pp.321-322, 2008 Fourth IEEE International Conference on eScience, 2008
www.biocep.net

[2] Simon Urbanek, Martin Theus (2003). iPLOTS, high interaction graphics for R. Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)
www.iplots.org