

Learning Math and Statistics on the Cloud,

Towards an EC2-based Google Docs-like portal for teaching/learning collaboratively with R and Scilab

Karim Chine
Cloud Era Ltd
Cambridge, UK
Karim.chine@polytechnique.org

Abstract—the Elastic-R portal allows educators and students to easily allocate, use and manage cloud resources and to work on the cloud, from standard web browsers, with the most commonly used statistical and mathematical environments such as R and Scilab. Elastic-R enables collaboration and resources sharing as well as interactive local or remote teaching sessions. Elastic-R is also a platform and provides frameworks and tools that make it easy for educators to create their own cloud-based e-Learning tools.

Cloud computing, R, Scilab, Virtual Research Environments, Virtual Learning Environments, remote teaching

I. INTRODUCTION

Cloud computing represents a new way to deploy computing technology in which dynamically scalable and virtualized resources are provided as a service over the Internet. Amazon Elastic Cloud (EC2) is an example of Infrastructure-as-a-Service (IaaS) that anyone can use today to access infinite computing capacity on demand within a sustainable environment that enables collaboration and resource sharing. This model of allocating processing power holds the promise of a revolution in scientific and statistical computing as well as in e-Learning. However, bringing new era for research and education still requires new software that bridges the gap between the scientists, educators and students' everyday tools and the cloud. The R statistical tool is a case in point. It is a free, application that attracts considerable use and method development in biostatistics, bioinformatics, genetics, etc. It is used for teaching statistics in most of the Universities all over the globe. Scilab has established itself as a credible alternative to Matlab and is widely used in applied mathematics courses.

II. R/SCILAB AND THE CLOUD

Elastic-R is a software platform, a Google docs-like portal and a workbench for data analysis that makes using R and Scilab on the cloud even simpler than using them locally. Elastic-R synergizes the usage scenarios of the R and the Scilab Scientific Computing Environments (SCEs) with the usage scenarios of the cloud and empowers them with what the cloud has best to offer:

a) User-friendly and flexible access to IaaS through the on-demand activation, composability and duplicability

of virtual appliances (machine, disks, ..).

b) Easy collaboration: sharable users' cloud resources are located on the same infrastructure and the virtual machine instances are publically accessible from anywhere.

c) On-demand elasticity. The cloud user can choose the capacity of his virtual machine instance (number of cores, memory size). The user can also choose to run any number of virtual machines and can shut them down anytime.

d) Applications deployment flexibility

e) Recording capabilities: The user can snapshot his/her virtual machines and virtual disks anytime and can archive those snapshots for later reuse or for sharing with other cloud users.

III. ELASTIC-R, AN OPEN PLATFORM FOR COMPUTING

The R and scilab environments capabilities can be extended thanks to the packages mechanism. Anyone can create his own R library or scilab module and use it in conjunction with all the existing toolkits. The created R package can be made accessible to the community very easily which explains the exponential growth of the R packages public repositories and the success of R. Nonetheless, R and scilab have many shortcomings and limitations. Elastic-R brings to the R and Scilab ecosystems the missing features and enables them to be applied in many more situations, in various different ways. By extending R's logic of openness and extensibility, Elastic-R builds an environment where all the artifacts and resources of computing become "pluggable". The Java or Ajax Elastic-R workbench allows the scientist, the educator and the student to easily assemble various synergetic capabilities:

a) The processing capability by connecting to a remote R/Scilab engine running at any location could it be a cloud, a grid or a local machine.

b) The mathematical and numerical capability by importing R/Scilab packages into the R/Scilab workspace

c) The orchestration capability by using the S language for programming with data, by using the various embedded scripting languages interpreters and by calling the different Elastic-R programming interfaces.

d) *The interaction capability*: by opening the various Elastic-R Workbench built-in views and by importing the Elastic-R client-side extensions (plugins).

e) *The persistence capability*: by setting the Elastic-R working directory to a local disk, an NFS server or a virtual disk on the cloud (Amazon Elastic Block Stores, EBS).

IV. VIRTUALIZATION AND E-LEARNING

Besides being free and mostly open source and therefore accessible to students and educators, Elastic-R provides education-friendly features that only proprietary software could offer so far (for example the centralized and controlled server-side deployment of the Scientific Computing Environments) and enables new scenarios and practices in the teaching of statistics and applied mathematics. With Elastic-R, it becomes possible for educators to hide the complexity of R, Scilab, Matlab, etc. with User Interfaces such as the Elastic-R plugins and spreadsheets. These are very easy to create and to distribute to students. The User Interfaces reduce the complexity of the learning environment and keep beginning students away from the steep learning curves of R, Scilab or Matlab. Once created by one educator, the User Interfaces can be shared, reused and improved by other educators. Dedicated repositories can be provided to centralize the efforts and contributions of the community of educators and help them sharing the insight gained in using this new environment. One could envisage these methods being used from primary schools to graduate-level studies.

Educators can adapt the Elastic-R virtual machines images to the specific needs of their courses and tutorials. For example, after choosing the most appropriate machine image, they can add to it the missing R packages, the required data files, install the missing tools, etc. The new image can then be provided to students on USB keys or made accessible on an IaaS-style cloud. In the first case, the students need only to have Java and a virtual machine player (the free VMware player for example) installed on their laptops to run the Elastic-R workbench and to connect to a computational engine on the virtual machine. In the second case, they need only a browser. Once again, a virtual machine prepared by one educator can be shared, reused and improved by other educators. The virtual machine is fully self-contained: the code needed to run the workbench or the plug-ins prepared by the educator can be delivered by the virtual appliance itself thanks to the Elastic-R code server that runs at startup. The interaction between the student and the SCE as well as the artifacts produced are saved within the Elastic-R-enabled-virtual machine. The educator can retrieve the USB keys used by the students (or connect to the virtual machine instance on the IaaS-style cloud) and checks not only the validity of the different intermediate results they obtained but also the path they followed to get those results.

The collaboration capabilities of the workbench open also new perspectives in distributed learning. The educator can connect anytime to the SCEs of students at any location. He can see and update their environments and guide them remotely. Collaborative problem solving becomes also possible and can be used as a support for learning.

The Elastic-R portal itself is available as a public virtual machine image (AMI) on Amazon Elastic Cloud. That AMI can be used to deploy a decentralized private virtual collaboration and learning environment. Because cloud resources are not free, students shouldn't use their own Amazon cloud accounts. Elastic-R provides a mechanism of secure digital tokens that can be delivered by the University IT manager to researchers, educators and students. The tokens allow them to start virtual machines for a specified number of hours and use them for their research or their courses.

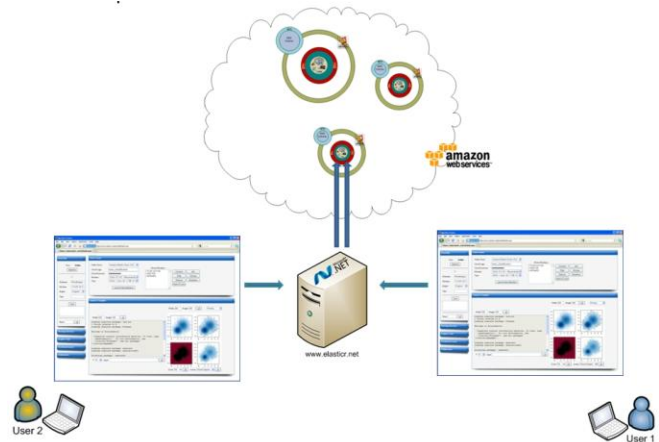


Figure 1. Elastic-R: Virtual e-Research/e-Learning environment

CONCLUSION

Elastic-R will enable educators to experiment new teaching methods. It may create an ecosystem for virtual e-Learning resources production and sharing and it demonstrates that the cloud is an invaluable new asset for e-Learning.

REFERENCES

- [1] R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- [2] <http://www.scilab.org>
- [3] www.elasticr.net/portal
- [4] Amazon, Inc., "Amazon Elastic Compute Cloud." [Online]. Available: aws.amazon.com/ec2
- [5] Mohammed Al-Zoube, "E-Learning on the Cloud", International Arab Journal of e-Technology, Vol. 1, No. 2, June 2009
- [6] Karim Chine, "Scientific Computing Environments in the age of virtualization, toward a universal platform for the Cloud" pp. 44-48, 2009 IEEE International Workshop on Opensource Software for Scientific Computation (OSSC), 2009

