

R and the Cloud

K. Chine ^a

^a Cloud Era Ltd
24 Langham Road,
Cambridge
UK

karim.chine@cloudera.co.uk

Mots clefs: Cloud Computing, EC2, e-Learning, distant education, e-Science, Collaboration, Science Gateways, Big data, Analytics-as-a-Service

Cloud Computing is holding the promise of democratizing access to computing infrastructures and deeply impacting research and education. However, the question "How will we bring the Infrastructure-as-a-Service paradigm to the data scientist's desk and to the statistics classroom?" has remained unanswered. The Elastic-R Software platform proposes new concepts and frameworks to address this question: R, Python, Matlab, Spreadsheets, etc. are made accessible as articulated, programmable and collaborative components within a virtual and immersive environment for scientific research and higher education.

Teachers can easily and autonomously prepare interactive R-based custom learning environments and share them like documents in Google Docs. They can use them in the classroom or remotely in a distant learning context. They can also associate them with on-line-courses. Students are granted seamless access to pre-prepared, controlled and traceable learning environments. They can share their R sessions to receive guidance from Teachers or to solve problems in collaboration. Costs may be hidden to the students by allowing them to access temporarily shared institution-owned resources or using tokens that a teacher can generate using institutional cloud accounts.

Scientists can easily use the cloud as a ubiquitous and scriptable collaborative environment for traceable and reproducible data analysis and computational research. The cloud becomes a user friendly Google-Docs-like platform where all the artifacts of computing can be produced by any number of geographically distributed real-time collaborators and can be stored, published and reused. Big data access and analysis are simplified and made accessible to wider range of research professional. Science Gateways (graphical user interfaces for data science; set of tools, applications, and data integrated via portals) are made R-scriptable and hence easy to create, publish and update on the fly from the R command line: their use becomes an intrinsic part of the process of programming with Data.

The presentation will give an overview of the synergies that exist between R and the state-of-the-art cloud technologies. Elastic-R on Amazon's public cloud will be demonstrated via real-world applications in education, in bioinformatics and in finance.

Références

- [1] Karim Chine (2010). Learning math and statistics on the cloud, towards an EC2-based Google Docs-like portal for teaching / learning collaboratively with R and Scilab, icalt, pp.752-753, 2010 10th IEEE International Conference on Advanced Learning Technologies.
- [2] Karim Chine (2010). Open science in the cloud: towards a universal platform for scientific and statistical computing. In: Furht B, Escalante A (eds) Handbook of cloud computing, Springer, USA, pp 453–474. ISBN 978-1-4419-6524-0
- [3] www.elastic-r.net
- [4] aws.amazon.com
- [5] www.coursera.org
- [6] sciencegateways.org