RosettaHUB, the Next Generation Data Science Platform

Fragmentation in the data science space reduces the productivity of data scientists and their ability to share, collaborate and make their results reproducible. Many powerful and high quality tools, languages and environments help tackle the different data analysis and scientific computing challenges but none of them covers on its own all the needs and requirements. One's PC is usually powerful enough to enable an important part of the problem solving scenarios but remote infrastructures are often needed for the flexibility they provide in scaling-up memory, compute power and storage. Those remote infrastructures, could they be public clouds, private clouds, clusters or supercomputers, are the only way to tackle big data challenges but they are also very heterogeneous and hard to use. The fragmentation also concerns data which is stored in a variety of ways and formats. Its access and manipulation often confronts data scientists with technical challenges.

Data scientists need an environment that significantly reduces those frictions and offers them a streamlined experience in their day-to-day interaction with tools, infrastructures, data and peers.

RosettaHUB delivers that missing environment and makes it possible for everyone to simultaneously use Python, R, Julia, SQL, Scala, Spark, Mathematica, ParaView, etc. as well as any computational code or library from within one single portal exposing web based and highly usable clouds management consoles, workbenches and notebooks. It offers a Google-docs like experience to data scientists and lets them use the different tools from anywhere and collaborate in real-time. It makes it possible to create and share resources on any infrastructure. It keeps track of all interactions with the environment and allows the reproduction of all the created artifacts.

Within rosettaHUB, cloud-enabled Python, R, Julia, Spark etc. share the same workspace and operate in the same memory, they can call each other seamlessly, share all their variables and access advanced collaborative visualization functionalities. The rosettaHUB environment is fully programmable and all its capabilities are accessible from desktop applications: The rosettaExcel spreadsheet allows data scientists to control seamlessly cloud resources as well as a plethora of data analysis tools expanding Excel's features with all the capabilities of those tools. A rosettaR package and a rosettaPy module bring the power of rosettaHUB to the R and Python local command lines, Software Developement Kits in many languages allow the use of rosettaHUB as a framework that empowers any existing applications or services.

RosettaHUB is also a social network, a ResearchGate-like environment for collaborative data science allowing researchers, teachers and students to create by themselves and use collaboratively virtual research and learning environments. It makes it very easy to create and distribute language-agnostic interactive and collaborative web applications and services: A dashboard or an interactive web application backed by any combination of cloud resources, data sets and code written in R, Python, SQL, Spark, etc. can be created in minutes and shared with specific users, groups or the general public. Those applications and services can be made available on the rosettaHUB marketplace. Following the github model, the free www.rosettahub.com portal will establish a new ecosystem and network that empowers every data scientist and helps democratize data science.