



The Elastic Analysis with Galaxy on the Cloud

Enis Afgan¹, Dannon Baker¹, Nate Coraor², Anton Nekrutenko², James Taylor¹

PENN STATE



EMORY

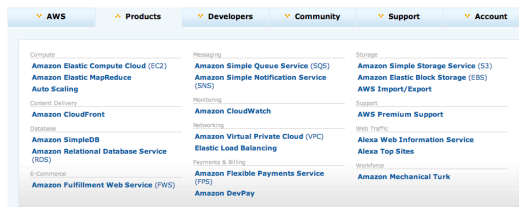
¹Department of Biology and Department of Mathematics & Computer Science, Emory University, Atlanta, GA, 30322, USA

²Huck Institutes of the Life Sciences and Department of Biochemistry and Molecular Biology, Pennsylvania State University, State College, PA, 16801, USA

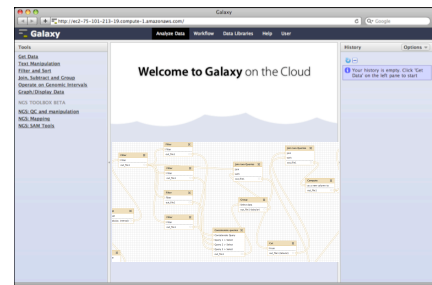
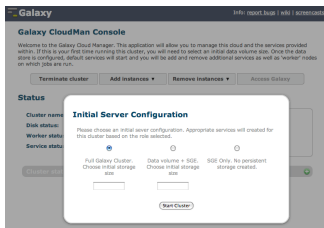
Abstract As experimental biologists become increasingly reliant on high-throughput data production, the scale and sophistication of computational infrastructure needed to support data storage and analysis has grown dramatically. In addition, the computational infrastructure needs to be coupled with the appropriate data analysis tools. Such an environment requires informatics support to setup, configure, and maintain the infrastructure. Moreover, once setup, the complete environment needs to be maintained during the periods of inactivity or low usage. For the experimentalists, such requirements represent a barrier to realizing the next step in science.

Cloud computing has recently emerged as a model that is well suited for the periodic computational requirements convenient to experimental biologists. However, **cloud computing resources are not yet suitable for immediate use by the experimentalists** because they still need to be configured and managed. To help in enabling seamless next-generation sequencing (NGS) analyses on the cloud, we have developed Galaxy CloudMan. Galaxy CloudMan is a comprehensive manager for running and managing cloud computing resources. **Cloud resources managed by Galaxy CloudMan are preconfigured with tools necessary for the NGS analyses.** Access and interaction with the preconfigured NGS tools is handled through Galaxy, an open-source, web based system that provides an integrated analysis environment where domain scientists can, without informatics expertise, interactively construct multi-step analyses, with outputs from one step feeding seamlessly to the next. Separate from the Galaxy analysis interface, **CloudMan offers a simple web based interface that allows anyone to acquire a desired number of computational and storage resources on a cloud infrastructure and access the familiar Galaxy interface and associated tools.** CloudMan automatically handles all aspects of resource acquisition, configuration, and data persistence, thus entirely insulating a user from the low-level computational details. With Galaxy CloudMan, an individual researcher can, without any informatics support, gain access to a complete NGS data analysis solution in a matter of minutes and release it once the analysis has completed, thus eliminating the need for the infrastructure maintenance.

Cloud computing out-of-the-box



CloudMan from the Galaxy Project



Why the cloud?

More tools

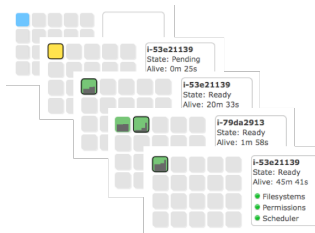


No need to share data

Choice clusters

High memory vs. High CPU

Less waiting



More storage



Behind CloudMan UI

