Ongoing efforts of a Galaxy solution for the BiGGrid & SARA HPC Cloud

Mattias de Hollander,[1] David van Enckevort,[2] Hailiang Mei,[2] Enis Afgan,[3] Eiko Kuramae[1]

[1]Microbial Ecology, Netherlands Institute of Ecology (NIOO/KNAW), Wageningen
[2]Bioassist, Netherlands Bioinformatics Centre (NBIC)
[3]Centre for Informatics and Computing, Ruđer Bošković Institute (RBI), Croatia

✉ m.dehollander@nioo.knaw.nl

Introduction

The rapid evolution of next generation sequencing technologies (NGS) together with decreasing costs are creating a challenge to store and analyze the vast amount of sequencing data that are generated by experimental biologists. Configuring suitable data analysis software and having access to readily available computation and storage resources are the two major bottlenecks faced by many research groups. In this on-going collaboration, NIOO and NBIC BioAssist program are building Galaxy@HPCcloud solution to jointly tackle these two challenges.

Advantages of Cloud Systems

→ Rapid elasticity (scale up/down dynamically)
→ Full adminstrational rights
→ Perfect for project-based research
→ Access to powerful compute systems

BiGGrid & SARA HPC Cloud (Calligo)

→ 19 x Intel Xeon 32 core processors (608 cores)
→ 19 x 256GB RAM (7.75 TB RAM)
→ 400TB shared storage

Setup Instructions

Details

→ Cloudman brings Galaxy to the Amazon Cloud
→ The HPC Cloud is powered by OpenNebula
→ A python module is written to support Cloudman
→ Ongoing project to support autoscaling and persistent storage

Outcomes

→ Public images to run Galaxy on the HPC Cloud
→ Instructions and source available on bitbucket and wiki
→ NBIC Galaxy server on the HPC Cloud with pre-installed tools
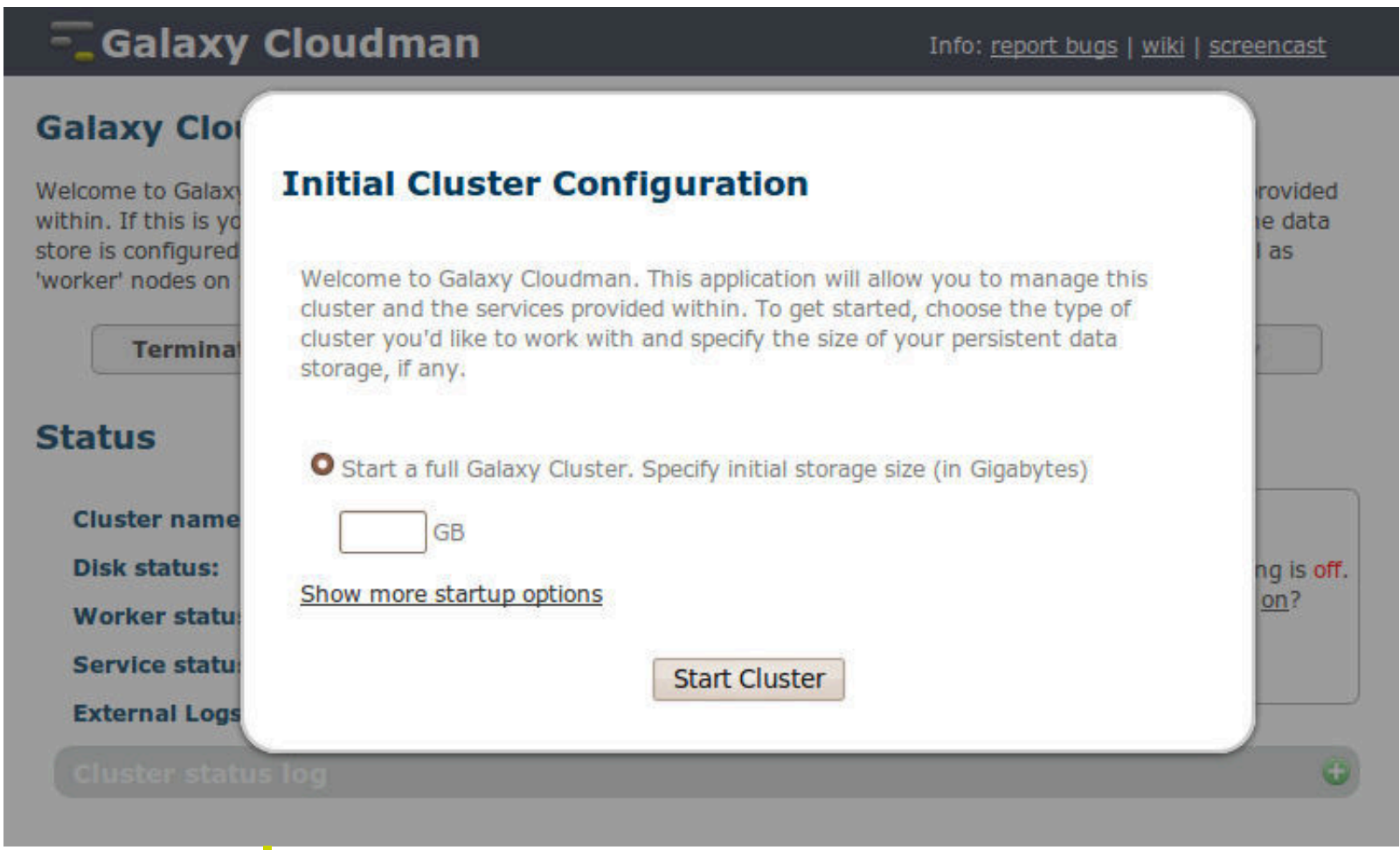
How it works
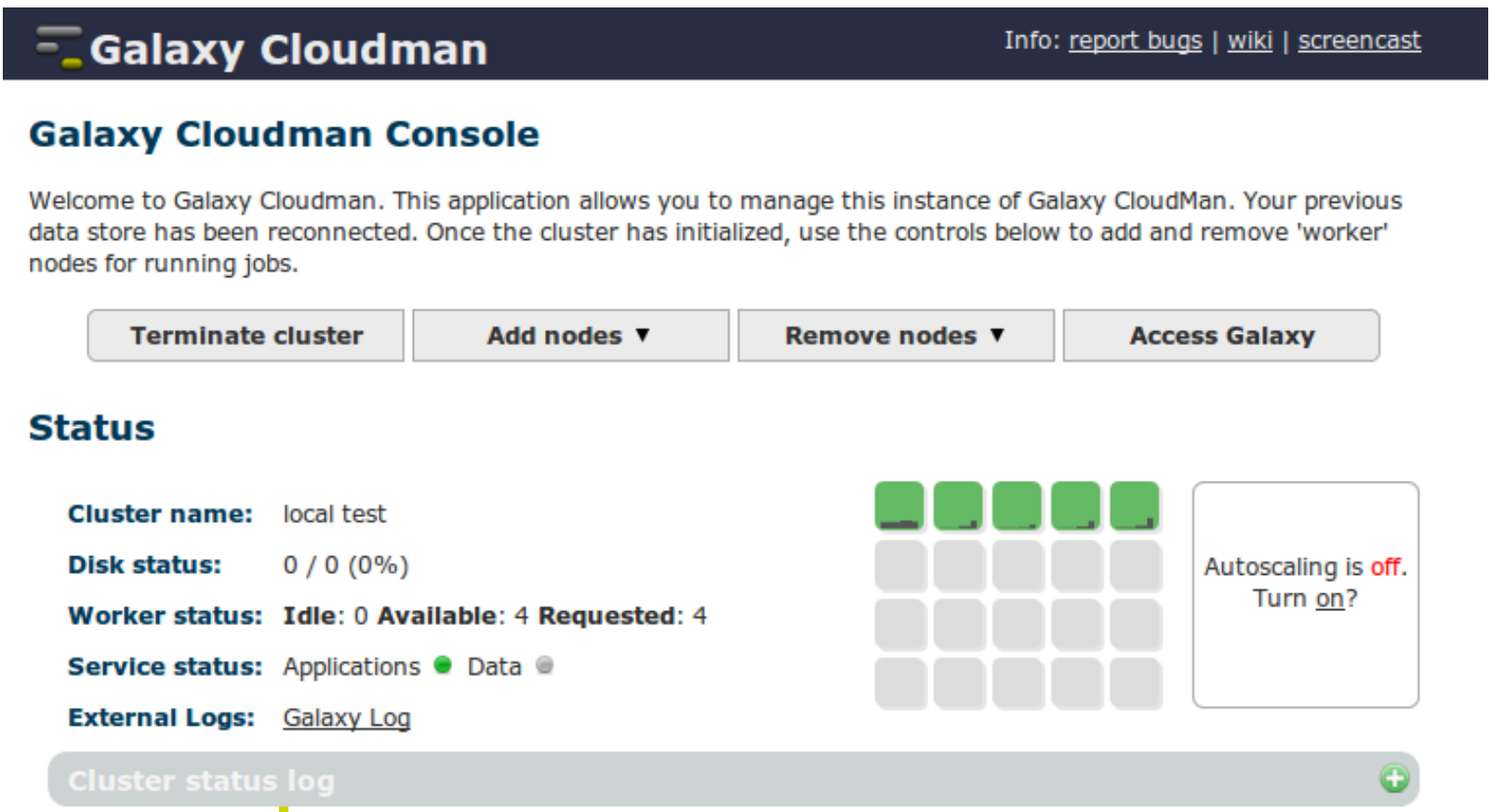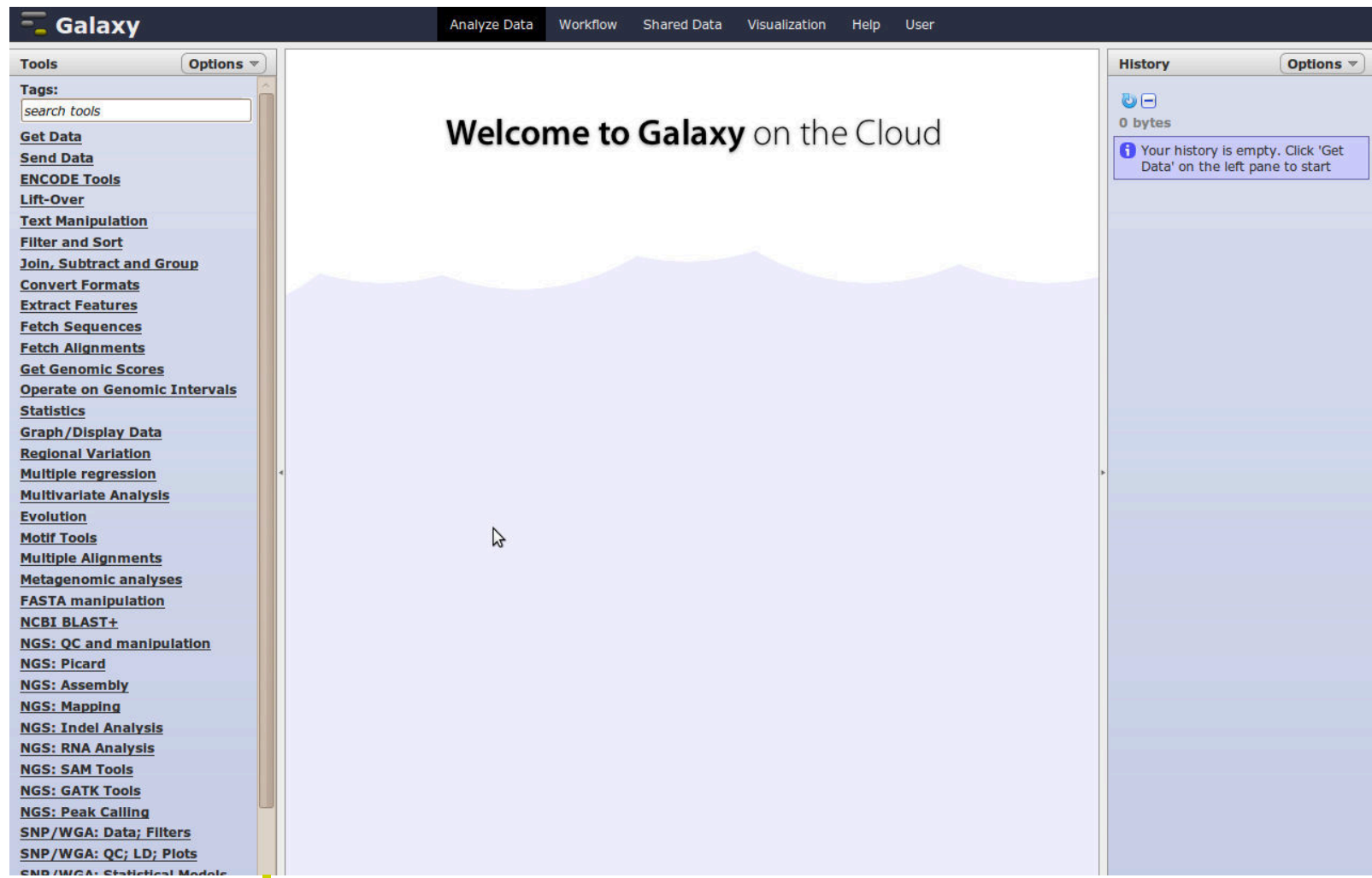

Figure 1 | Launching a Galaxy Instance

Figure 2 | Master and worker instances

Figure 3 | The Galaxy interface on the Cloud