

Providing Bioinformatics Services on Cloud

Tuesday, 9 April 2013 11:40 (25 minutes)

Summary

Improvements of experimental technologies forces biologists to face a deluge of data that require relevant tools and sufficient resources to be analyzed. The cloud helps bioinformatics experts to define virtual appliances with pre-installed tools and workflows, and helps scientists to deploy them, on demand, on national research infrastructures.

URL

<http://idee-b.ibcp.fr>

Impact

The platform IDB provides cloud services to the bioinformatics community with the StratusLab framework and our own developments. We have developed several bioinformatics appliances, predefined virtual machines, with installed and configured bioinformatics tools. Two examples of them are the 'biocompute' appliance (with BLAST, ClustalW2, Clustal-Omega, FastA, etc.) and the 'biodata' appliance (with databases Swiss-Prot, PROSITE, etc.) providing easy access to common tools and data.

The adoption of clouds for bioinformatics applications will be strongly correlated to the capability of cloud infrastructures to provide ease-of-use of common bioinformatics tools and access to reference biological databases. In that way, clouds for bioinformatics have to be connected with public bioinformatics infrastructures and this work is done in collaboration with the French Bioinformatics Network RENABI (www.renabi.fr) to help fulfill the requirements from the Bioinformatics community.

Description

Biologists and bioinformaticians simultaneously use many of the bioinformatics tools from the arsenal of thousands available from the international community. Most of the time, they also need to combine multiple software packages to study their data with public or their own analysis pipelines. For intensive usage, they have access to computing clusters through command line interfaces. The typical usage, however, is through web portals and services for the ease of use and for the capacity to compose these tools into pipelines. For several years, the focus has been put on providing such composable services and defining standards. But these bioinformatics applications can process gigabytes of data stored in flat-file databases like UNIPROT, EMBL or PDBseq. And they require access to reference databases in a POSIX manner (like NFS) to the cloud storage containing the biological data.

As identifying and running the relevant bioinformatics appliances could be difficult, we have developed a custom bioinformatics web interface to the cloud. This portal is coupled with the StratusLab Marketplace where the bioinformatics tools are referenced and tagged with RDF metadata. These metadata can be used to select the right bioinformatics appliance according to the desired tools (BLAST, ClustalW, etc.) or the kind of analysis to perform (sequence or structural analysis, assembling/mapping, etc.). The bioinformatics portal aims to make the cloud easy to use by non-computing and non-cloud-specialist scientists, like biologists and bioinformaticians. The available features allow the creation and termination of the virtual machines, the management of the persistent disks, and provide assistance for bioinformatics appliance selection and instance contextualization.

Primary authors: BLANCHET, Christophe (IDB IBCP FR3302); Mr GAUTHEY, Clément (IDB CNRS-IBCP)

Presenter: BLANCHET, Christophe (IDB IBCP FR3302)

Session Classification: Cloud Platforms

Track Classification: Cloud Platforms (Track Lead: M Drescher and M Turilli)