

CI-Dash: speeding up cloud computing in bioinformatics

Author : Muniba Faiza

Categories : [Cloud Computing](#), [Tools](#)

Date : November 25, 2015



After a lot of work in the field of bioinformatics, many of the living organisms' genome has been sequenced and a lot of information has been generated at RNA and protein level. This has given rise to a huge amounts of biological data whose storage is a issue now a days, because such an enormous data cannot be stored on a personal computer or on a local server. For this purpose cloud computing, a practice to manage, and process data by using remote servers hosted on internet has been introduced in bioinformatics, though the origin of cloud computing is not very clear.

CI-dash is a tool which facilitates research of novel bioinformatics data using Hadoop – a software that stores huge amount of data and provide a very easy access to that data in a relatively lesser time. This tool has been developed by Paul Hodor, Amandeep Chawla, Andrew Clark and Lauren Neal from Booz Allen Hamilton, USA.

The tool is “ci-dash”, it is a starter kit, which configures and apply the new hadoop clusters in a few minutes. It is provided by AWS (Amazon Web Services).

According to a paper published in Bioinformatics (Nov, 2015), **cl-dash is based on the distributed file system and MapReduce programming pattern.** Hadoop MapReduce is a software framework for easily writing applications which process vast amounts of data in-parallel on large clusters of personal computers or hardwares. With the help of cl-dash, a user can create clusters (or nodes which stores huge amount of data) as an 'admin', through a set of command line tools, which begins with 'cl-' (hence the name: 'cl dash'). A YAML configuration file (config.yml) is required to make a new cluster can be created in minutes. Once the Hadoop cluster is formed, the user can easily access the data.

Such tools are required for further storage space requirement because biological data is increasing, thereby, the demand for large data storage space is also required. cl-dash has provided a good pathway for managing such a huge data.

NOTE:

An exhaustive list of references for this article is available with the author and is available on personal request, for more details write to muniba@bioinformaticsreview.com.

Sharing is caring. Spread the love!

- [Print](#)
- [Email](#)
- [LinkedIn](#)
- [Twitter](#)
- [Facebook](#)
- [Google](#)
-