Concerns and Challenges of Cloud Platforms for Bioinformatics

B

Nicoletta Dessì
Università degli Studi di Cagliari, Italy

Barbara Pes
Università degli Studi di Cagliari, Italy

INTRODUCTION

In recent years, computer advances have played an important role in promoting scientific research in biological areas such as genomics, proteomics and other "-omic" subfields which rely heavily on suitable computational infrastructures for managing large-scale data. In particular, the flood of data from genome sequences has given rise to "bioinformatics", an interdisciplinary research domain which employs a wide range of computational techniques derived from scientific disciplines (such as statistics, machine learning, applied mathematics etc..) for managing biological data. To get to understand the current application fields of bioinformatics, it is necessary to consider the following aspects.

A first aspect is about the massive production and spread of biological data around the web. Generated within a short period of time and stored in a growing number of web resources, the increasing amount of biological data has introduced new challenges about its management and exploitation.

For example, thanks to next generation sequencing instruments and ICT advances, areas of life sciences that were previously distant from each other (in the ideology, analysis practices, toolkits etc.) are now able to share and analyze data in transparent and reproducible fashion. This interdisciplinary research task calls for the integration of information with multiple levels of granularities from several web resources that often represent information and data in different ways.

In this respect, bioinformatics increasingly deals with providing technical approaches to support interdisciplinary scientific knowledge which relies on working with concepts from different areas in constant evolution and more and more requires experimental techniques, scientific approaches and collaborative management of data (Bosin, Dessì, & Pes, 2007).

A second aspect is about recent advances in computer science that significantly influence the development of computational tools in bioinformatics. Specifically, the service-oriented paradigm has provided a new way of thinking biological resources in terms of computational infrastructures by positioning services as primary functional elements for data integration. Several biomedical organizations (such as the National Center for Biotechnology Information (NCBI) and The National Center for Biomedical Ontology (NCBO)) provide web portals that expose Web services for searching data. Existing techniques for web content classification, search, and visualization seem to be actually inadequate to satisfy the biologist's needs because accessing these heterogeneous systems from the Internet is not straightforward without the availability of standard and common interfaces.

In this respect, bioinformatics research is devoted to search explicit and automatic ways of joining information to improve the usability of web resources.

Finally, the rapid development of the Internet has provided an opportunity to investigate about

the use of state-of-the-art technology for the construction of a new generation of tools that integrate plain data sources, public program-mable APIs and any kind of available services. Usually referred to as Web2.0 applications, these tools rely on open APIs or reusable services. The availability of biomedical ontologies dramatically increases the range of benefits and the usages derived from these applications that often support a deeper analysis of data by taking into account the semantic information (Dessì, Pascariello, & Pes, 2014).

Considering the aforementioned aspects, it is clear that bioinformatics research addresses three main challenges namely:

1. Storing and analyzing large amount of het-erogeneous data.
2. Enabling knowledge extraction from several web resources and collaboration through user-friendly interfaces.
3. Promoting solutions for offering different categories of services to end-users.

Nowadays, the cloud computing paradigm represents a primary solution to these challenges as it extends the role of the Internet to enable a new form of distributed system for large-scale data processing.

As significant component of the Web 2.0 world, Cloud computing expresses a service-based archi-tecture where computing services are delivered on-demand to customers over a network in a self-service fashion that is independent of device and its location. This innovative technology enables the introduction of new computational paradigms that could have a great influence on facing the above challenges. The application of cloud computing in bioinformatics is still preliminary, although and increasing number of biological applications is relying on the Cloud for processing large datasets (Calabrese & Cannataro, 2016; Schadt, Linder-man, Sorenson, Lee, & Nolan, 2011). Conversely, few work has been done to explore the use of ar-chitectural cloud-based solutions (Mrozek, 2015).

This chapter aims investigating the extent to which cloud technology offers a viable platform for developing and deploying applications that support users in searching and integrating in-formation offered by bioinformatics resources. The chapter outlines the basic features that such computing applications should exhibit and the challenging issues they deal with. The architecture and the functionality of the cloud-based environ-ments are presented to stress how cloud platforms could offer added-value service components and flexibility that make their adoption attractive for bioinformatics.

The chapter also sketches some application scenarios that illustrate the potential of cloud-based environments for integrating information from multiple web resources, including biomedi-cal portals.

BACKGROUND ON CLOUD COMPUTING

Cloud technology endows users with a service-based environment that provides hardware and system components and refers to software ap-plications delivered as services over the Internet (Lin, Fu, Zhu, & Dasmalchi, 2009).

Cloud resources are usually organized into bottom-up layers, depending on the type of service:

-   **Infrastructure-as-a-Service (IaaS):** Provides suitable hardware, including physical servers and network bandwidth services directly to end users which do not own enough processing power, storage capacity, applications to meet their needs. These resources can be 'rented' from a cloud provider on an as-needed basis and accessed from anywhere via an Internet connection. Cloud services employ a me-tering system that divides the computing resources in appropriate blocks and users pay for the service as an operating expense without incurring any significant initial capital costs.

Related Content

Mechanical Transmission Model and Numerical Simulation Based on Machine Learning
Pan Zhang (2023). *International Journal of Information Technologies and Systems Approach (pp. 1-15).*
www.irma-international.org/article/mechanical-transmission-model-and-numerical-simulation-based-on-machine-learning/318457

The Challenges of Teaching and Learning Software Programming to Novice Students
Seyed Reza Shahamiri (2018). *Encyclopedia of Information Science and Technology, Fourth Edition (pp. 7392-7398).*
www.irma-international.org/chapter/the-challenges-of-teaching-and-learning-software-programming-to-novice-students/184437

An Integrated Approach to Supply Chain Simulation
Nenad Stefanovicand Bozidar Radenkovic (2018). *Encyclopedia of Information Science and Technology, Fourth Edition (pp. 5398-5410).*
www.irma-international.org/chapter/an-integrated-approach-to-supply-chain-simulation/184243

An Enhanced Text-Classification-Based Arabic Information Retrieval System
Sameh Ghwanmeh, Ghassan Kannan, Riyad Al-Shalabiand Ahmad Ababneh (2009). *Utilizing Information Technology Systems Across Disciplines: Advancements in the Application of Computer Science (pp. 37-44).*
www.irma-international.org/chapter/enhanced-text-classification-based-arabic/30715

Educational Ontology Development
Galip Kayaand Arif Altun (2018). *Encyclopedia of Information Science and Technology, Fourth Edition (pp. 1441-1450).*
www.irma-international.org/chapter/educational-ontology-development/183859