

Chapter 4

Concerns and Challenges of Cloud Platforms for Bioinformatics

Nicoletta Dessì

Università degli Studi di Cagliari, Italy

Barbara Pes

Università degli Studi di Cagliari, Italy

ABSTRACT

Bioinformatics traditionally concerns applying computational approaches for the management and the exploitation of large volumes of biomedical data that continues to expand in size and in distribution. Although the application of cloud computing in biomedical areas is still preliminary, an increasing number of biomedical applications rely on the Cloud for processing large datasets. This chapter investigates the extent to which cloud technology offers a viable platform for developing and deploying applications that support users in searching and integrating information 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 environments are presented to stress how cloud platforms could offer added-value service components and flexibility that make their adoption attractive for bioinformatics.

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.

DOI: 10.4018/978-1-5225-7489-7.ch004

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 programmable 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 heterogeneous 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 architecture 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

9 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/concerns-and-challenges-of-cloud-platforms-for-bioinformatics/213582

Related Content

Use of Technology in Problem-Based Learning in Health Science

Indu Singh, Avinash Reddy Kundurand Yun-Mi Nguy (2019). *Advanced Methodologies and Technologies in Medicine and Healthcare* (pp. 365-375).

www.irma-international.org/chapter/use-of-technology-in-problem-based-learning-in-health-science/213612

A Comprehensive Analysis of Findings on Clinical, Radiological, and Histopathological Correlations in Cholesteatoma

Vikas S. Kulkarni, Abhay D. Havle, Anagha Y. Rajguruand Agus Sudaryanto (2024). *Advancements in Clinical Medicine* (pp. 275-286).

www.irma-international.org/chapter/a-comprehensive-analysis-of-findings-on-clinical-radiological-and-histopathological-correlations-in-cholesteatoma/346206

Multi-Criteria Decision-Making Techniques for Histopathological Image Classification

Revathi T., Saroja S., Haseena S.and Blessa Binolin Pepsi M. (2019). *Histopathological Image Analysis in Medical Decision Making* (pp. 103-138).

www.irma-international.org/chapter/multi-criteria-decision-making-techniques-for-histopathological-image-classification/212541

Pediatric Visual Acuity Testing

Gayathri Srinivasan (2022). *The Pediatric Eye Exam Quick Reference Guide: Office and Emergency Room Procedures* (pp. 44-66).

www.irma-international.org/chapter/pediatric-visual-acuity-testing/296160

Building Gene Networks by Analyzing Gene Expression Profiles

Crescenzo Gallo (2019). *Advanced Methodologies and Technologies in Medicine and Healthcare* (pp. 27-44).

www.irma-international.org/chapter/building-gene-networks-by-analyzing-gene-expression-profiles/213581