

Chapter 12

Cloud-Based Computing Architectures for Solving Hot Issues in Structural Bioinformatics

Dariusz Mrozek

Silesian University of Technology, Poland

ABSTRACT

Bioinformatics as a scientific domain develops tools that enable understanding the wealth of information hidden in huge volumes of biological data. However, there are several problems in bioinformatics that, although already solved or at least equipped with promising algorithms, still require huge computing power in order to be completed in a reasonable time. Cloud computing responds to these demands. This chapter shows several cloud-based computing architectures for solving hot issues in structural bioinformatics, such as protein structure similarity searching or 3D protein structure prediction. Presented architectures have been implemented in Microsoft Azure public cloud and tested in several projects developed by Cloud4Proteins research group.

1. INTRODUCTION

Bioinformatics is an interdisciplinary scientific domain, which develops algorithms, methods and software tools for discovering knowledge that is hidden in the wealth of biological data. For example, bioinformatics tools used for comparison of two genes, genomes or proteins help in understanding of evolutionary relationships between two organisms. Such information is not apparent at first glance when looking at raw data and requires appropriate software tools and algorithms in order to become visible. Bioinformatics helps in understanding genetic basis of unique diseases, like Alzheimer's, cancer, diabetes mellitus, providing information on mutations in nucleotide sequences of DNA or RNA strands. Based on the information scientists can distinguish types of particular diseases and medical doctors are able to adjust their therapies to the type of disease. Knowledge of protein structures and other biologi-

DOI: 10.4018/978-1-5225-8903-7.ch012

cal molecules is a foundation for the creation of new drugs, which translates into better quality of life. Therefore, it is not surprising that bioinformatics has become a part of many areas of biology, medicine and modern pharmacy.

Structural bioinformatics is the branch of bioinformatics, which focuses on the analysis, comparison and modelling of the three-dimensional structures of proteins, DNA and RNA acids, and other molecules and molecular complexes. This sub-discipline of bioinformatics has emerged as a result of rapid increase in the number of three-dimensional macromolecular structures available in databases such as the Protein Data Bank (<http://www.pdb.org>) (Berman et al., 2000). Structural bioinformatics provides general purpose methods for processing information about biological macromolecules with aim to create new knowledge and a high-resolution understanding of biology. It relies on the belief that having this high-resolution structural information about biological systems will allow scientists to precisely reason about the function of these systems and the effects of modifications or perturbations (Gu & Bourne, 2009). However, performing analyses on such detailed level requires increased computational resources that are sometimes available only in big data centers. Bioinformatics is developing very dynamically in recent years and the development is also supported and accelerated by development of new computing architectures and new computing models. The concept of Cloud computing became one of the elements catalyzing the development of the domain.

Following the definition of NIST, Cloud computing is “a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction” (Mell & Grance, 2011). In practice, cloud computing allows to run applications and services on a distributed network using virtualized system and its resources, and at the same time, allows to abstract away from the implementation details of the system itself. The use of cloud platforms can be particularly beneficial for companies and institutions that need to quickly gain access to a computer system which has a higher than average computing power. In this case, the use of cloud computing services can be more cost effective and faster in implementation than using the owned resources (servers and computing clusters) or buying new ones. For this reason, cloud computing is widely used in business. However, the concept of cloud computing is also becoming increasingly popular in scientific applications for which theoretically infinite resources of the cloud allow to solve the computationally intensive problems. This chapter shows several, tested on Microsoft Azure platform, cloud architecture patterns for solving hot issues in structural bioinformatics, including protein structure similarity searching or 3D protein structure prediction. Presented architectures have been implemented and tested in several projects developed by Cloud4Proteins research group.

2. BACKGROUND

Today's bioinformatics is struggling with many problems and use a variety of approaches to solve them. This chapter provides broad definitions of the problems, which Cloud4Proteins group tries to solve, and focuses on technologies that are used to achieve the solutions.

20 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/cloud-based-computing-architectures-for-solving-hot-issues-in-structural-bioinformatics/228628

Related Content

Fungi-Mediated Detoxification of Heavy Metals

Suchhanda Ghosh (2021). *Recent Advancements in Bioremediation of Metal Contaminants* (pp. 205-219).

www.irma-international.org/chapter/fungi-mediated-detoxification-of-heavy-metals/259573

Dynamical Spectra in Two-Dimensional Dusty Plasmas

Aamir Shahzad, Zakia Rafiq, Alina Manzoor and Muhammad Kashif (2022). *Emerging Developments and Applications of Low Temperature Plasma* (pp. 34-48).

www.irma-international.org/chapter/dynamical-spectra-in-two-dimensional-dusty-plasmas/294709

Biosynthesis of Silver Nanoparticles for Study of Their Antimicrobial Effect on Plasma-Treated Textiles: Silver Coating of Plasma-Treated Cotton Fabric

Shazia Shukrullah, Muhammad Anwar, Muhammad Yasin Nazam and Inzamam Ul Haq (2022). *Emerging Developments and Applications of Low Temperature Plasma* (pp. 149-166).

www.irma-international.org/chapter/biosynthesis-of-silver-nanoparticles-for-study-of-their-antimicrobial-effect-on-plasma-treated-textiles/294715

Innovation and Intellectual Property Rights: The Case of Soybean Seeds in Argentina and the United States

Mercedes Campi (2019). *Biotechnology: Concepts, Methodologies, Tools, and Applications* (pp. 1944-1965).

www.irma-international.org/chapter/innovation-and-intellectual-property-rights/228699

Biosorption of Heavy Metals: Biological Approach to Control the Industrial Waste

Neelesh Babu, Vinay Mohan Pathak, Akash Akash and Navneet Navneet (2019). *Biotechnology: Concepts, Methodologies, Tools, and Applications* (pp. 1898-1909).

www.irma-international.org/chapter/biosorption-of-heavy-metals/228697