International Journal of Advanced Computer and Mathematical Sciences.
Vol 1, Issue 1, Dec, 2010, pp 1-6. http://bipublication.com

CLOUD COMPUTING APPLICATIONS IN COMPUTATIONAL SCIENCE

Manisha B. Jadhav, * Vishnu J. Gaikwad, C. V. Patil, G. S. Deshpande
Department of Computer Science and IT, MGM College of CS and IT Nanded 431605, MS, India
*Department of Biotechnology, Tatyasaheb Kore Institute of Engineering and Technology, Warananagar,
Tal: Panhala, Dist: Kolhapur, Pin: 4116113, MS, India

Abstract

Cloud computing is a style of computing in which dynamically scalable and often virtualized resources are provided as a service. Users need not have knowledge of, expertise in, or control over the technology infrastructure in the "cloud" that supports them. and all its explanation. We are on the verge of a new era in computing, in which innovative software running on intelligent devices ("clients"), complemented by Internet-based data storage and services (the "cloud"), will offer users more engaging, seamless experiences across their computers, cell phones and other devices.
Many services in bioinformatics take longer than just a few seconds to complete, for example querying over large or multiple databases and computationally intensive calculations like protein structure prediction, identification of protein domains, families and functional sites, and sequence alignments.

This next generation of computing holds enormous potential to stimulate economic growth and enable governments to reduce costs, increase transparency and expand services tocitizens.
KEYWORDS: Cloud computing, economic growth, bioinformatics, protein domains, alignments.

INTRODUCTION:

Cloudcomputing is Internetbased computing, whereby shared resources, software and information are provided to computers and other devices on-demand, like a public utility.

It is a paradigm shift following the shift from mainframe to client-server that preceded it in the early '80s. Details are abstracted from the users who no longer have need of, expertise in, or control over the technology infrastructure "in the cloud" that supports them.[1] Cloud computing describes a new supplement, consumption and delivery model for IT services based on the Internet, and it typically involves the provision of dynamically scalable and often virtualized resources as a service over the Internet.[2, 3] It is a byproduct and consequence, of the ease-of-access to remote computing sites provided by the Internet.[4]

The term cloud is used as a metaphor for the Internet, based on the cloud drawing used in the past to represent the telephone network [5], and later to depict the Internet in computer network diagrams as an abstraction of the underlying infrastructure it represents.[6] Typical cloud computing providers deliver common business applications online which are accessed from another web service or software like a web browser, while the software and data are stored on servers.

A technical definition is "a computing capability that provides an abstraction between the computing resource and its underlying technical architecture (e.g., servers, storage, networks), enabling convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction."[7] This definition states that clouds have five essential characteristics: on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service.[7]

Since 2007, the number of trademark filings covering cloud computing brands, goods and services has increased at an almost exponential rate.[8] As companies sought to better position themselves for cloud computing branding and marketing efforts, cloud computing trademark filings increased by 483% between 2008 and 2009.[9] In 2009, 116 cloud computing trademarks were filed, and trademark analysts predict that over 500 such marks could be filed during 2010.[10]

According to Nicholas Carr, the strategic importance of information technology is diminishing as it becomes standardized and less expensive. He argues that the cloud computing paradigm shift is similar to the displacement of electricity generators by electricity grids early in the 20th century.[11]

Other factors impacting the scale of any potential cost savings include the efficiency of a company's data center as compared to the cloud vendor's, the company's existing operating costs, the level of adoption of cloud computing, and the type of functionality being hosted in the cloud.[12, 13]

Applicational Reviews:

Peer-to-Peer (P2P) computing allows peer nodes (computers) to share content directly with one another in a decentralized manner. In pure P2P computing, there is no notion of clients or servers since all peer nodes are equal and concurrently be both clients and servers. The goals of P2P computing include cost sharing or reduction, resource aggregation and interoperability, improved scalability and reliability, increased autonomy, anonymity or privacy, dynamism, and ad-hoc communication and collaboration [14].

Market-oriented computing views computing resources in economic terms such that resource users will need to pay resource providers for utilizing the computing resources [15].

Industry analysts have made bullish projections on how Cloud computing will transform the entire computing industry. According to a recent Merrill Lynch research note [16], Cloud computing is expected to be a "$160-billion addressable market opportunity, including $95-billion in business and productivity applications, and another $65-billion in online advertising". Another research study by Morgan Stanley [17] has also identified Cloud computing as one of the prominent technology trends.

As the computing industry shifts toward providing Platform as a Service (PaaS) and Software as a Service (SaaS) for consumers and enterprises to access on demand regardless of time andlocation, there will be an increase in the number of Cloud platforms available. Recently, several academic and industrial organizations have started investigating and developing technologies and infrastructure for Cloud Computing. Academic efforts include Virtual Workspaces [18], OpenNebula [19], and Reservoir [20].

Amazon Elastic Compute Cloud (EC2) [21] provides a virtual computing environment that enables a user to run Linux-based applications. The user can either create a new Amazon Machine Image (AMI) containing the applications, libraries, data and associated configuration settings, or select from a library of globally available AMIs.

The user then needs to upload the created or selected AMIs to Amazon Simple Storage Service (S3), before he can start, stop, and monitor instances of the uploaded AMIs. Amazon EC2 charges the user for the time when the instance is alive, while Amazon S3 [22] charges for any data transfer (both upload and download). Google App Engine [23] allows a user to run web applications written using the Python programming language.

Microsoft Azure [24] aims to provide an integrated development, hosting, and control Cloud computing environment so that software developers can easily create, host, manage, and scale both Web and non-web applications through Microsoft data centers. To achieve this aim, Microsoft Azure supports a comprehensive collection of proprietary development tools and protocols which consists of Live Services, Microsoft .NET Services, Microsoft SQL Services, Microsoft SharePoint Services, and Microsoft Dynamics CRM Services. Microsoft Azure also supports Web APIs such as SOAP and REST to allow software developers to interface between Microsoft or non-Microsoft tools and technologies.

Sun network.com (Sun Grid) [25] enables the user to run Solaris OS, Java, C, C++, and FORTRAN based applications. First, the user has to build and debug his applications and runtime scripts in a local development environment that is configured to be similar to that on the Sun Grid. Then, he needs to create a bundled zip archive (containing all the related scripts, libraries, executable binaries and input data) and upload it to Sun Grid. Finally, he can execute and monitor the application using the Sun Grid web portal or API. After the completion of the application, the user will need to download the execution results to his local development environment for viewing.

Cloud Services for Bio- and Cheminformatics

XMPP services were developed using xws4j for bioinformatics and cheminformatics. A service, HIVPred, was set up that predicts the susceptibility of sequences for seven known HIV protease inhibitors using a proteochemometric model [26].

Many services in bioinformatics take longer than just a few seconds to complete, for example querying over large or multiple databases, and computationally intensive calculations like protein structure prediction [27], identification of protein domains, families and functional sites [28], and sequence alignments [29].

The CloudBurst project [30] explored the benefits of using Hadoop as a platform for alignment of short reads. Cloud-Burst is capable of reporting all alignments for millions of human short reads in minutes, but does not scale well to human resequencing applications involving billions of reads. Whereas CloudBurst aligns about 1 million short reads per minute on a 24-core cluster, a typical human resequencing project generates billions of reads, requiring more than 100

days of cluster time or a much larger cluster.

Hadoop-based software tool that combines the speed of the short read aligner Bowtie [31] with the accuracy of the SNP caller SOAPsnp [32] to perform alignment and SNP detection for multiple whole-human datasets per day.

Architecture

based options and fewer IT skills are required for implementation (in-house).[37]

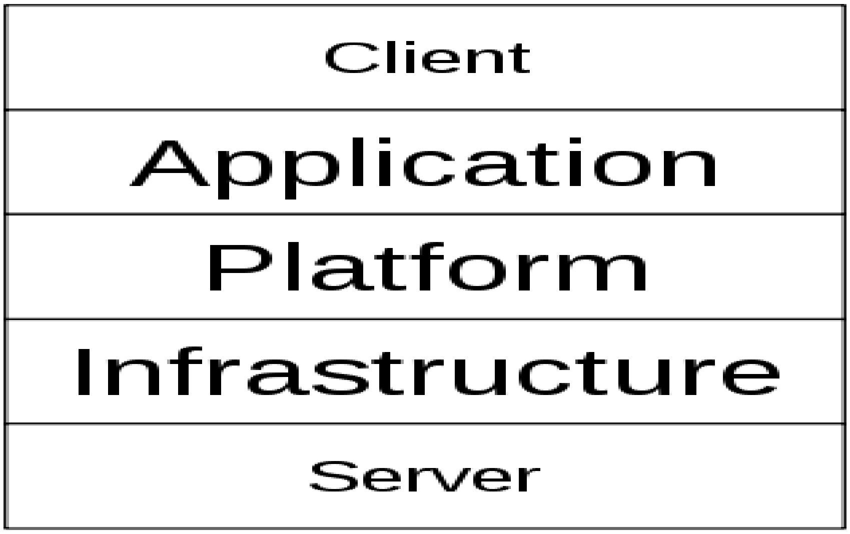

Fig1. Architecture of cloud computing

Cloud architecture,[33] the systems architecture of the software systems involved in the delivery of cloud computing, typically involves multiple cloud components communicating with each other over application programming interfaces, usually web.[34]
This resembles the Unix philosophy of having multiple programs each doing one thing well and working together over universal interfaces. Complexity is controlled and the resulting systems are more manageable than their monolithic counterparts.

Key Features:

Agility improves with users' ability to rapidly and inexpensively re-provision technological infrastructure resources.[35]

Cost is claimed to be greatly reduced and capital expenditure is converted to operational expenditure [36]. This ostensibly lowers barriers to entry, as infrastructure is typically provided by a third-party and does not need to be purchased for one-time or infrequent intensive computing tasks. Pricing on a utility computing basis is fine-grained with usage-based options and fewer IT skills are required for implementation (in-house).[37]

Device and location independence [38] enable users to access systems using a web browser regardless of their location or what device they are using (e.g., PC, mobile). As infrastructure is off-site (typically provided by a third-party) and accessed via the Internet, users can connect from anywhere.[37]

Multi-tenancy enables sharing of resources and costs across a large pool of users thus allowing for:

Centralization of infrastructure in locations with lower costs (such as real estate, electricity, etc.), **Peak-load capacity** increases (users need not engineer for highest possible load-levels)

Maintenance cloud computing applications are easier to maintain, since they don't have to be installed on each user's computer. They are easier to support and to improve since the changes reach the clients instantly.
Although cloud computing is often assumed to be a form of "green computing", there is as of yet no published study to substantiate this assumption. [39]

Conclusion:

As organizations cope with a dynamically changing business environment, IT managers look to cloud computing as a means to maintain a flexible and scalable IT infrastructure that enables business agility.

In June 2009, F5 Networks conducted a study examining the adoption of cloud computing by enterprise IT managers. The study found that although significant confusion regarding the definition of the cloud exists, IT managers are aggressively deploying cloud computing initiatives to accomplish business objectives. Additionally, the study found that widespread enterprise adoption is contingent upon solving access, security and performance concerns.

References:

1. Distinguishing Cloud Computing from Utility Computing
2. Gartner Says Cloud Computing Will Be As Influential As E-business
3. Gruman, Galen (2008-04-07). "What cloud computing really means". InfoWorld. Retrieved 2009-06-02.
4. "Cloud Computing: Clash of the clouds". The Economist. 2009-10-15. Retrieved 2009-11-03.
5. http://www.sellsbrothers.com/writing/intro2tapi/default.aspx?content=pstn.htm
6. The Internet Cloud
7. Cloud Computing Definition, National Insitute of Standards and Technology, Version 15
8. NSA has massive database of Americans' phone calls
9. Feature Guide: Amazon EC2 Availability Zones
10. FISMA compliance for federal cloud computing on the horizon in 2010
11. Nicholas Carr on 'The Big Switch' to cloud computing
12. 1 Midsize Organization Busts 5 Cloud Computing Myths
13. Cloud Computing Savings - Real or Imaginary?
14. D. S. Milojicic, V. Kalogeraki, R. Lukose, K. Nagaraja, J. Pruyne, B. Richard, S. Rollins, and Z. Xu. Peer-to-Peer Computing. Technical Report HPL-2002-57R1, HP Laboratories, Palo Alto, USA, 3 July 2003.
15. D. Abramson, R. Buyya, and J. Giddy. A Computational Economy for Grid Computing and its Implementation in the Nimrod-G Resource Broker, Future Generation Computer Systems,18(8):1061-1074, Oct. 2002.
16. D. Hamilton. 'Cloud computing' seen as next wave for technology investors. Financial Post, 4 June 2008.
17. http://www.financialpost.com/money/story.html?id=562877 [18 July 2008]
18. Morgan Stanley. Technology Trends. 12 June2008.http://www.morganstanley.com/institutional/techresearch/pdfs/TechTrends062008.pdf [18 July 2008]
19. K. Keahey, I. Foster, T. Freeman, and X. Zhang. Virtual workspaces: Achieving quality of service andquality of life in the Grid. Scientific Programming, 13(4):265-275, October 2005.
20. OpenNebula Project. http://www.opennebula.org/ [23 July 2008]
21. Reservoir Project. http://www.reservoir-fp7.eu/ [30 Oct. 2008]
22. Amazon Elastic Compute Cloud (EC2). http://www.amazon.com/ec2/ [18 July 2008]
23. Amazon Simple Storage Service (S3). http://www.amazon.com/s3/ [18 July 2008]
24. Google App Engine. http://appengine.google.com [18 July 2008]
25. Microsoft Azure. http://www.microsoft.com/azure/ [30 Oct. 2008]
26. Lapins M, Eklund M, Spjuth O, Prusis P, Wikberg JES: Proteochemometric modeling of HIV protease susceptibility. BMC Bioinformatics2008, 9:181.
27. Wallner B, Larsson P, Elofsson A: Pcons.net: protein structure prediction meta server. Nucleic Acids Res 2007, 35:W369-74.
28. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, Finn RD, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Laugraud A, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, Mistry J, Mitchell A, Mulder N, Natale D, Orengo C, Quinn AF, Selengut JD, Sigrist CJA, Thimma M, Thomas PD, Valentin F, Wilson D, Wu CH, Yeats C: InterPro: the integrative protein signature database. Nucleic Acids Res 2009, 37:D211-5.
29. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG: Clustal W and Clustal X version 2.0. Bioinformatics 2007, 23:2947-2948.

30. Schatz MC: CloudBurst: highly sensitive read mapping withMapReduce. Bioinformatics 2009, 25:1363-1369.
31. Langmead B, Trapnell C, Pop M, Salzberg SL: Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 2009, 10:R25
32. Li R, Li Y, Fang X, Yang H, Wang J, Kristiansen K: SNP detectionfor massively parallel whole-genome resequencing. GenomeRes 2009, 19:1124-1132.
33. Building GrepTheWeb in the Cloud, Part 1: Cloud Architectures
34. Cloud Maturity is Accelerating: More Than Just Reaction To The Hype?
35. Infrastructure Agility: Cloud Computing as a Best Practice
36. Recession Is Good For Cloud Computing – Microsoft Agrees
37. Defining “Cloud Services” and “Cloud Computing”
38. The new geek chic: Data centers
39. James Urquhart (January 7, 2010). "Cloud computing's green paradox". CNET News. Retrieved March 12, 2010. "...there is some significant evidence that the cloud is encouraging more compute consumption