

Aggregate Service Using One Stop Shop and Web Crawler in Cloud Computing

Dr. Puspita Dash¹, Ajeeth Kumar R², Arivazhagan R², Karthikeyan V²

¹Assistant Professor, Department of Information Technology, Sri Manakula Vinayagar Engineering College, Puducherry, India

²Department of Information Technology, Sri Manakula Vinayagar Engineering College, Puducherry, India

ABSTRACT

Cloud computing services has become an important paradigm as it is reliable and provides a cost effective way of storing and hosting applications. Cloud storage is growing exponentially and monitor the data in a secure manner. Cloud computing is the delivery of different services through the Internet. These resources include tools and applications like data storage, servers, databases, networking, and software. As long as an electronic device has access to the web, it has access to the data and the software programs to run it. The one-stop shop expanded over time to include business services. The shifted from a wide product offering to capture more of the customer's grocery purchase to one of offering all the complementary products and services to a client in a particular area. Web crawlers are mainly used to create a copy of all the visited pages for later processing by a search engine, that will index the downloaded pages to provide fast searches.

Keyword – Cloud Computing, Cloud Storage, Web Crawler, Customer's Grocery.

Article Info

Volume 9, Issue 3

Page Number : 638-641

Publication Issue

May-June-2022

Article History

Accepted : 03 June 2022

Published : 20 June 2022

I. INTRODUCTION

cloud computing is the delivery of computing services including servers, storage, databases, networking, software, analytics, and intelligence over the Internet to offer faster innovation, flexible resources, and economies of scale. You typically pay only for cloud services you use, helping lower your operating costs, run your infrastructure more efficiently and scale as your business needs change Cloud Computing is a general term for anything that involves delivering

hosted services over the Internet. Instead of a static system architecture, Cloud Computing supports the ability to dynamically scale up and quickly scale down, offering cloud consumers high reliability, quick response times, and the flexibility to handle traffic fluctuations and demand. 7 Cloud Computing also supports multi tenancy, providing systems configured in such a way that they can be pooled to be shared by many organizations or individuals. Virtualization technology allows cloud vendors to convert one server into many virtual machines, thereby

eliminating client-server computing with single-purpose systems. This maximizes hardware capacity and allows customers to leverage economies of scale.

In [1], address the author could see nwhat is happening in each scope, knowing the development of knowledge in each range and saw the results of research conducted in each field in[3], the author has proposed a The technical details of the proposed system can be classified into two groups including the customer services and the restaurant management services. The customer services consist of the waiting queue module, the food ordering module, the payment module and the customer satisfaction survey module. The restaurant management services consist of the food management module and the staff management module.

A one - stop shop is a firm that offers a multitude of products or services to its customers, all under one roof, so to speak. A one-stop shop can refer to a literal roof a specific physical location where all the business a client has can be carried out or it can refer to a company that handles a variety of goods or services.

A web crawler downloads and indexes content from all over the Internet. The goal of such a bot is to learn what every webpage on the web is about, so that the information can be retrieved when it's needed. Web crawlers copy pages for processing by a search engine, which indexes the downloaded pages so that users can search more efficiently. Crawlers consume resources on visited systems and often visit sites unprompted. Issues of schedule, load, and "politeness".

II. WEB CRAWLER

A web crawler downloads and indexes content from all over the Internet. The goal of such a bot is to learn what every webpage on the web is about, so that the information can be retrieved when it's needed. Web crawlers copy pages for processing by a search engine, which indexes the downloaded pages so that users can search more efficiently. Crawlers consume resources

on visited systems and often visit sites unprompted. Issues of schedule, load, and "politeness".

Web search engines and some other websites use Web crawling or spidering software to update their web content or indices of other sites' web content. Web crawlers copy pages for processing by a search engine, which indexes the downloaded pages so that users can search more efficiently.

The work proposed in [6] , explores the first is that the heuristic search alogrithm designed in this paper is not suitable for some binding websites that do not pay attention to the development of norms. The page structure of this kind of special website is very different from the features set extraction of website is relatively low In [7] ,a web crawler works as a simple indexer which can extract all the links which contains the word or string given by a user. There is no limit for te st cases. It will search for all possible cases. Hence it will result in giving many results at a time; some of which are not useful. Thus, this crawler takes more time for searching for all possible combinations and takes more memory in saving all those links. The work submitted in [8],the author proposes a design takes big data of popular science books consumption as an example to carry out data mining. After data capture, data visulaization analysis and mining are carried out. It mainly analyze a overall sale of popular science books, the correlation between the price and sales of popular , the correlation between sales ranking and sales, the correlation between price and comment number In the idea focused in [9] , the author has proposed The average number keywords in a domain web page may range from 500 to 2000. Fetching these keywords ids from these Keyword 14 Service would normally take 100 to 200ms.The fetching time from the unordered_map is almostconstant even after a huge amount of keywords are insertedinto the structure. In [10], the paper The processing flow of the crawler controller is as follows. First, receive the keywords entered by the user and the specified number of pages to be queried; Second, start multiple threads to search for relevant content

by keywords from Google search engine, Baidu search engine, Zhihu, etc., and get results from Get Pages Module; Third, call the Parser Module for page parsing; Fourth, obtain the target uniform resource locator (URL) queue through the get URLs module and the URLs deduplication module

III. PROPOSED SYSTEM

There are some limitations in search services for getting a information about a particular word or other related data. The following diagram in the figure 1 gives the proposed system which is used to enhance a data or an information of search service.

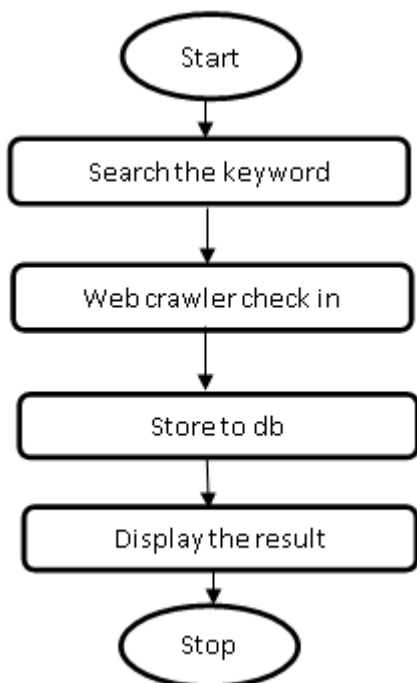


Fig.1 Block diagram of the current system

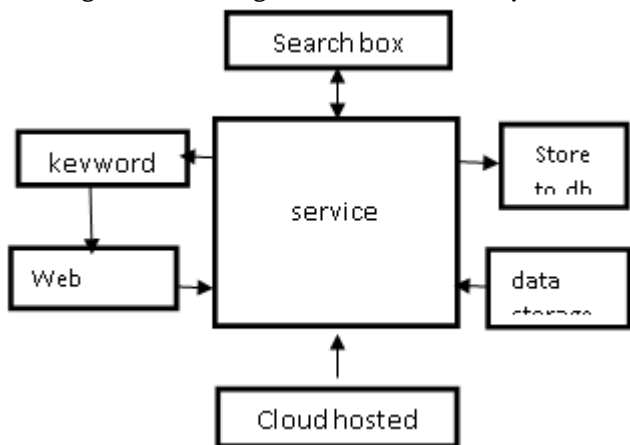


Fig.2 Control flow diagram

The service works as a request and respond methodology, it is a cloud hosted service and it can be implemented in many other services.

Search service can be accessed from any devices by the user and search for a information. The word or a sentence searched by user is sent to web crawler and database. The web crawler which will take a keyword from a searched word and search accordingly, the crawler arrange the order store it in index format for quick identification of the data.

The word searched by the user will be stored in database(db). It is used for quick access propose if we search a same word twice. It will check with the database if the searched word is new it will be added to the database. In here we use two storage of database. One database is for holding the information and data for the user search and the other database is for storing the word which is searched by the user, in each way the service should be utilized by the user.

While searching for a keyword the request message will be passed to one to another node for processing the task.

Algorithm used in searching the content or word in search service is Breath first search.

Breadth-first search is a graph traversal algorithm that starts traversing the graph from the root node and explores all the neighboring nodes. Then, it selects the nearest node and explores all the unexplored nodes. While using BFS for traversal, any node in the graph can be considered as the root node.

It starts at the tree root and explores all nodes at the present depth prior to moving on to the nodes at the next depth level. Extra memory, usually a queue, is needed to keep track of the child nodes that were encountered but not yet explored.

BFS can be used to find the neighboring locations from a given source location. In a peer-to-peer network, BFS algorithm can be used as a traversal method to find all the neighboring nodes. Most torrent clients, such as BitTorrent, uTorrent, etc. employ this process to find "seeds" and "peers" in the network. BFS can be used in web crawlers to create

web page indexes. It is one of the main algorithms that can be used to index web pages. It starts traversing from the source page and follows the links associated with the page. Here, every web page is considered as a node in the graph.

BFS and its application in finding connected components of graphs

IV. CONCLUSION

Cloud Computing is used to manage a data and also support hosting. Cloud hosting makes applications and websites accessible using cloud resources. Search service is most widely used in many application and services to derive the user need. The search service is very reliable and utilized by everyone. In this way our project “Aggregate service using web crawler and one stop shop in cloud computing ” will help the people those who need desired data or the information it will fetched them with exact content about everything.

V. REFERENCES

- [1]. Shan Liu, Keming Yue ,Lu Liu and Ting Guo , “Research on Saas Model Based on Cloud Computing” pages (1959 - 1962) , Information Management,Communicates,Electronic and Automation Control Conference(IMCEC 2018)
- [2]. Aobing Sun and Jialin Zhou ,“ Cloud Services ” , International Conference on Cloud and service Computing, pages (309 - 314), 2018
- [3]. Tarek Menouer, Amina Khedimi, Chirstophe Cerin and Mohammed Chahbar, “Scheduling Service Function Chains with Dependencies in the Cloud”, pages (1 - 3), International Conference on cloud networking 2020
- [4]. Kitsiri Chochiang, Pkaypreak Ung and Narongrit Bunsaman, “ One stop Restaurant Service Application”, 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON) 2020 , pages(482 - 485)
- [5]. Muhammad Mahreza Maulana, Arif Imam Suroso,” Smart Governance for One-Stop Shop Services of Bio-Business ”, International Conference on Computer Science and Its Application in Agriculture 2020,pages (1- 6)
- [6]. Widjajarto, et al. “Architecture Model of Information Technology Infrastructure Based on Service Quality at Government Institution.” *Procedia Computer Science*, vol. 161, Elsevier B.V., 2019, pp. (841–50)
- [7]. Naresh Kumar and Manjeet Singh ,“Framework for Distributed Semantic Web Crawler”, International Conference on Computational Intelligence and Communication Networks 2015 ,pages (1403 – 1407)
- [8]. Hui Liu, Kunpeng Wang and Zikai Liu, “Application and research of heuristic search algorithm in crawler field”, 2nd International Conference on Information Technology and Computer Application – 2020, pages (41 - 44)
- [9]. Swasti Singhal, Riya Sharma, Laxmi Ahuja and Ajay Rana,” Approach in Web Application for Regression Testing Using Crawlers – 2020 ,pages(263 - 266)
- [10]. Yuxin Wang, Qi Zhang , “Data Minig of Popular science books based on web crawler”, International Conference on Big Data, AI and IoT 2020, pages (92 - 95)
- [11]. Ragavan, Yesubai Rubavathi and John Singh ,”Crawler Framework for Category Search Engine”, International Conference on Emerging Trends in Infromation Technology 2020 , pages(1 - 5)
- [12]. Jiangying Xu, Lixin Du, Chunsun Duan and Mingyue Li ,” Application of Web Crawler in City Image Research”, 5th International Conference on Cloud Computing and Big Data Analytics 2020, pages 270 – 274

Cite this Article

Dr. Puspita Dash, Ajeeth Kumar R, Arivazhagan R, Karthikeyan V, "Aggregate Service Using One Stop Shop and Web Crawler in Cloud Computing", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 9 Issue 3, pp. 638-641, May-June 2022. Available at doi : <https://doi.org/10.32628/IJSRST2293134>
Journal URL : <https://ijsrst.com/IJSRST2293134>