

Visualization of Enterprise Grade Cloud Architecture for the Most Cost-Effective Application Deployment using Mining Tool

¹Dr.Senbagavalli.M, Associate professor, Alliance University, Bangalore, India.

²Dr.Vinodha.K, Associate professor, Alliance University, Bangalore, India.

³Dr.Jesudas.T, professor, Mahendra Engineering College, Namakkal, India.

ABSTRACT - Research and assessment by other companies and researchers is insufficient or ineffective in selecting the best provider of cloud to implement. We were therefore determined to expand our studies to include a solid, cost-effective implementation solution for all significant cloud platform providers. The manual process of collecting raw data from various social media platforms and SEO tools is a very tedious process and requires many skilled forces to complete it successfully. Yet it might contain lots of human errors which are common. Thus, to save time and money, we proposed an Architecture which required to work on data mining and validation tools to perform web scraping and reviewing of data, which yielded results within a short period of time.

Keywords: Cloud Computing, Visualization, Mining, Amazon Web Services, Google Cloud Platform (GCP).

I. INTRODUCTION

The industry has risen enormously in the public cloud computer. Unlike personal clouds used externally by organizations, public cloud systems are available to nearly anybody with a credit card. In a few minutes, clients in the public cloud can quickly turn dozens or even hundreds of VMs; they only pay for what they really use, without massive upfront investment. This flexibility motivates more and more companies to migrate their apps to public-cloud systems, including governments, academia, businesses and content providers. In the wake of the trend, public cloud computing providers like Amazon Web Services (AWS), Google Cloud Platform (GCP), Microsoft Azure, IBM Bluemix, Alibaba cloud and Oracle cloud are increasingly being implemented by businesses. While the range of public cloud vendors encourages healthy competition, selecting the finest cloud is a challenging process for an implementation.

Imagine moving your email service to a cloud. It is simple to identify suppliers with the characteristics to help the request, since most have comparable set of characteristics. However, it is hard to determine which cloud provides the highest email processing efficiency, primarily because there is no detailed, extensive cloud provider efficiency comparison. Providers generally have unclear efficiency requirements (e.g., "every cloud server receives four virtual CPUs"). Many concentrates on virtual machine's computation speed and ignore other elements, such as the efficiency of storage and network facilities.

We have lately carried out a thorough analysis of the results of public cloud suppliers. This research leads to a systematic assessment of which cloud efficiency measures

best characterize. We can rapidly select the most efficient supplier for our implementation to properly adhere to our application requirements. Another issue is that the entry of data into excel sheets is a time-consuming method which plagues companies of all dimensions. The entered information manually can lead to common errors without adequate consideration. These errors might appear insignificant but may cost thousands of bucks to fix by your company.

We have written python scripts which detect these errors and inform data entry staff of errors they have made to fix them fast.

There is certainly a huge hype for cloud computing today, and many professionals see it as the future of information technology. McKinsey says that, with a need-oriented and optimized use of infrastructural assets, business companies in specific can reach a profit capacity of between 30 percent and 40%. There is a lengthy way to go for this exploitation before it is depleted. For instance, it is not well recognized that business companies are particularly suitable with the assistance of Open Source funds to construct and handle their own cloud alternatives. Platforms in this sector offer cost-effective, autonomous, innovative and versatile alternatives that are generally very easy to implement. Therefore, the comparison of such cloud platform providers is worth examining in depth.

Research and evaluation by other businesses or people are inadequate or ineffective to choose the finest cloud provider for our implementation. Thus, we were determined to extend our research in depth on all major cloud platform providers to come up with a robust and cost-effective deployment solution.

A. Related Work

Junjie Peng, a professor with the School of Computer Science at Shanghai University, along with his team of researchers have worked on a paper [1] that discusses about the comparison of different public cloud platforms based on the pros and cons. But the paper details about unpopular cloud platforms and doesn't discuss in any form about the price estimations of these platforms and their services. Another researcher Mr. Srikanth Kandula, principle researcher at Microsoft, along with his team have worked on a research project [2] that details about the comparison of the public cloud platforms based on their performance metrics. Yet again this research work is limited to only summary of differences only on the performance difference but not on their price estimations nor the services provided.

Some researchers from California, USA have worked on a detailed research work on the comparison of efficiency and cost of public cloud platforms [3]. But this research is also limited to only Amazon Web Services (AWS) and VMWare. There already exists a chrome extension for Facebook data miner, but the features of this tool only include text data such as reviews, posts and details about a page. The limitations of this tool are that it cannot be used for mining photos and videos, and also the data extracted is saved into separate excel sheets which becomes a complex process to integrate again into a consolidated file.

In this paper [9], proposed an Architecture and design of enterprise-based application by considering cost and availability parameters. They limited with number of parameters and security issues. In Paper [10], they have discussed about cloud computing services, characteristics and Applications and It helps many enterprises and any individuals understand how cloud computing can provide them with reliable, customized and cost-effective services in various Applications. This research is limited with security and privacy issues continue to be the biggest biggest concern on cloud.

II. CLOUD SYSTEM DESIGN

A. Deployment Methods

When individuals start debating about cloud computing, they usually mention about one of four feasible application code deployment solutions: Infrastructure as a Service (IaaS), Container as a Service (CaaS), Platform as a Service (PaaS), or Software as a Service (SaaS).

i. Infrastructure as a Service (IaaS)

In an IaaS model, a cloud provider houses traditionally in-house infrastructure components such as servers, network and storage hardware, virtualization or a hypervisor layer. The IaaS provider also provides several services to support these components of the infrastructure. Detailed billing, monitoring, log access, security, load balancing and clustering and storage resilience, like backup, replication and recovery are all available. These are increasingly

policies-based services, enabling IaaS users for important infrastructures to achieve a greater degree of automation and orchestration. For instance, a user may use policies to balance loads to maintain availability and performance of applications.

ii. Container as a Service (CaaS)

Containers as a Service (CaaS) is a cloud service that allows container-based virtualization to upload, organize, run, scale, manage, and stop containers by software developers and IT departments. A CaaS provider usually provides a framework for users to use the service. The applications programming interface (API) or a web interface is typically employed by providers. CaaS falls between Infrastructure as a Service (IaaS) and Platform as a Service (PaaS) within the range of cloud computing services. CaaS is, however, best used as an IaaS subset. A container instead of a virtual Machine (VM) or a hardware hosting system that is traditionally used to support IaaS environments is the basic resource for CaaS.

iii. Platform as a Service (PaaS)

PaaS typically does not replace the whole IT infrastructure of a company. The company relies instead on PaaS vendors, such as application hosting or Java development, for key services. A PaaS provider creates a robust and optimized environment for users to install apps and data sets. Instead of building and retaining the fundamental infrastructure and facilities, users can concentrate on generating and operating apps. Many PaaS brands focus on the growth of apps. These systems provide a computer and transport infrastructure as well as text editing, version control, compilation and evaluation utilities to make fresh software more rapidly and effectively available to designers. A PaaS item can also allow growth groups, irrespective of their physical place, to operate together.

iv. Software as a Service (SaaS)

SaaS is closely associated with the provider of application service (ASP) and delivery models of computer software on demand. The SaaS host application management model is like ASP, where the provider hosts the software and provides it through the internet to approved end-users. The provider gives clients network-based access to a single copy of an application, which has been specifically created for the distribution of SaaS, in the SaaS model on-demand software. The source code of the application is the same for all clients, which are handed out to all clients when fresh characteristics or characteristics are implemented. The client data for each model can be stored locally, in the cloud or at the local level as well as in the cloud, depending on the service level agreement (SLA). SaaS applications may be integrated by organizations using other software programming interfaces (APIs). For example, a company can use the SaaS Provider APIs to integrate such tools with the SaaS offer and create its own software tools.

B. System Architecture

The applications developed were all standalone applications which did not need any server and client connection.

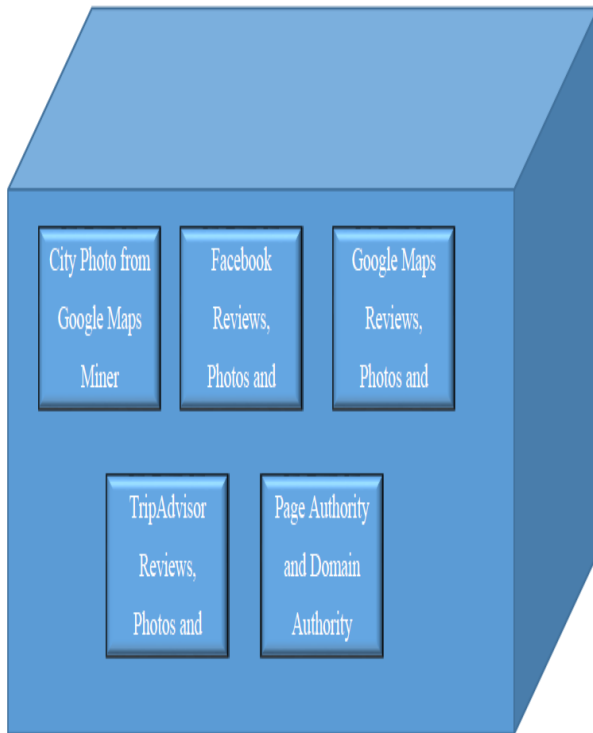


Fig.1 Standalone Client Application

i. Development Environment

The development environment consisted of Windows 10 operating system, with Intel Core i3 Quadcore processor and 4GB RAM. My code was written using Python IDLE editor and executed with the Python 3.5.1rc1 release via command prompt.

ii. Deployment Environment

The applications were deployed to Azure Cloud Virtual Machine (B2MS machine type), running on Windows Server 2012 consisting of 2 vCPUs, 8GB RAM and 1TB Premium SSD.

C. Implementation and results

All the scripts in this section use Selenium library for automating the scripts and openpyxl library to read and write into excel files. The requests module was used to send and retrieve HTTP GET requests.

i. City Photos from Google Maps Mining Script

This script loads a given input dataset of city names along with the state they are present in. Each city is searched on Google maps and the respective photos of the searched city is mined and loaded into the local system.

ii. Facebook Reviews, Photos and Videos Mining Script

This script loads an excel sheet containing a dataset of Facebook page URLs. Each Facebook page is loaded, and

the respective reviews, photos and videos are extracted and loaded into the local system.

iii. Google Maps Reviews, Photos and Videos Mining Script

This script loads an excel sheet containing a dataset of Google Place ID. A Google PlaceID is a text ID that uniquely defines a specific location. The script then reverse engineers this PlaceID to extract the address of every location and opens the respective Google Maps page. From the respective pages reviews, photos and videos are extracted and loaded into the local system.

iv. TripAdvisor Reviews, Photos and Videos Mining Script

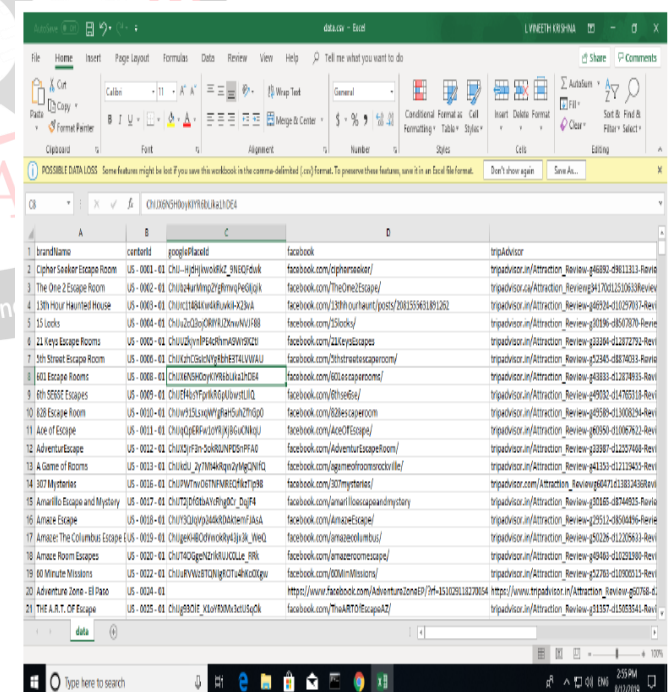
This script loads an excel sheet containing a dataset of TripAdvisor page URLs. Each TripAdvisor page is loaded, and the respective reviews, photos and videos are extracted and loaded into the local system.

v. Page Authority and Domain Authority Mining Script

This script loads an excel sheet containing a dataset of multiple URLs. Each URL is loaded, and the respective page is opened in selenium along with Mozbar chrome extension loaded. The data generated by the Mozbar tool is extracted by the script and stored into an excel sheet in the local system.

Results

i. Screenshots



	A	B	C	D	E
1	brandName	centart	google	facebook	tripadvisor
2	Cyber Sankar Escape Room	US-0001-01	ChUjyHw0k2_9N0C4uik	facebook.com/cyberankar	tripadvisor.com/Attraction_Review-g46891-d811133-Review
3	The One 2 Escape Room	US-0002-01	ChUjyHw0k2_9N0C4uik	facebook.com/TheOne2Escape	tripadvisor.com/Attraction_Review-g46891-d811133-Review
4	13th Hour Haunted House	US-0003-01	ChUjyHw0k2_9N0C4uik	facebook.com/13thHourHauntedHouse	tripadvisor.com/Attraction_Review-g46891-d811133-Review
5	15 Locks	US-0004-01	ChUjyHw0k2_9N0C4uik	facebook.com/15Locks	tripadvisor.com/Attraction_Review-g46891-d811133-Review
6	21 Key Escape Rooms	US-0005-01	ChUjyHw0k2_9N0C4uik	facebook.com/21KeyEscapeRooms	tripadvisor.com/Attraction_Review-g46891-d811133-Review
7	5th Street Escape Room	US-0006-01	ChUjyHw0k2_9N0C4uik	facebook.com/5thStreetEscapeRoom	tripadvisor.com/Attraction_Review-g46891-d811133-Review
8	601 Escape Rooms	US-0008-01	ChUjyHw0k2_9N0C4uik	facebook.com/601EscapeRooms	tripadvisor.com/Attraction_Review-g46891-d811133-Review
9	6th Street Escape Rooms	US-0009-01	ChUjyHw0k2_9N0C4uik	facebook.com/6thStreetEscapeRooms	tripadvisor.com/Attraction_Review-g46891-d811133-Review
10	628 Escape Room	US-0010-01	ChUjyHw0k2_9N0C4uik	facebook.com/628EscapeRoom	tripadvisor.com/Attraction_Review-g46891-d811133-Review
11	Ace of Escape	US-0011-01	ChUjyHw0k2_9N0C4uik	facebook.com/AceOfEscape	tripadvisor.com/Attraction_Review-g46891-d811133-Review
12	Adventure Escape	US-0012-01	ChUjyHw0k2_9N0C4uik	facebook.com/AdventureEscapeRoom	tripadvisor.com/Attraction_Review-g46891-d811133-Review
13	A Game of Rooms	US-0013-01	ChUjyHw0k2_9N0C4uik	facebook.com/Agamesofroomsrockville	tripadvisor.com/Attraction_Review-g46891-d811133-Review
14	307 Myrtles	US-0015-01	ChUjyHw0k2_9N0C4uik	facebook.com/307Myrtles	tripadvisor.com/Attraction_Review-g46891-d811133-Review
15	Amazilla Escape and Mystery	US-0017-01	ChUjyHw0k2_9N0C4uik	facebook.com/jamallscapandmystery	tripadvisor.com/Attraction_Review-g46891-d811133-Review
16	Amaze Escape	US-0018-01	ChUjyHw0k2_9N0C4uik	facebook.com/AmazeEscape	tripadvisor.com/Attraction_Review-g46891-d811133-Review
17	Amaze: The Columbus Escape	US-0019-01	ChUjyHw0k2_9N0C4uik	facebook.com/amazecolumbus	tripadvisor.com/Attraction_Review-g46891-d811133-Review
18	Amaze Room Escapes	US-0020-01	ChUjyHw0k2_9N0C4uik	facebook.com/amazeroomescapes	tripadvisor.com/Attraction_Review-g46891-d811133-Review
19	60 Adventure Missions - El Paso	US-0022-01	ChUjyHw0k2_9N0C4uik	facebook.com/60AdventureMissions	tripadvisor.com/Attraction_Review-g46891-d811133-Review
20	The A.B.T. Of Escape	US-0024-01	ChUjyHw0k2_9N0C4uik	facebook.com/TheABTOfEscape	tripadvisor.com/Attraction_Review-g46891-d811133-Review

Fig. 2 Sample Input File containing Social Media Links

Fig.3 Screenshot of Facebook Miner Tool in Action

Fig.4 Screenshot of Output Generated by Facebook Miner Tool

Fig. 5 Screenshot of TripAdvisor Miner Tool in Action

Fig. 6 Screenshot of City Photo Miner Tool in Action

III. DATA VALIDATION & PROCESSING

Implementation and results

i. Data Validation

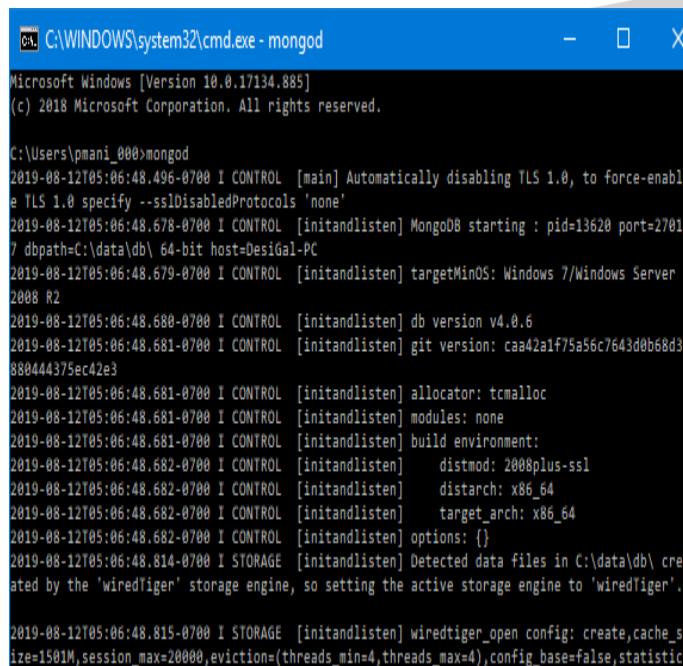
This script loads a consolidated excel sheet which contains raw data provided by the Data Mining team. It then iterates through every column to check if the data present in valid format provided by the company or not. If

the column is not present, we also need to check the comments section for the reasons for empty column using scripting itself and if there are no comments related to the column and the column is found invalid, the respective rows are then shifted to a new excel file called Invalid Data. If all the columns in a row are valid, only those rows are shifted to a new excel file called Validated Data.

ii. Data Processing

This script is used to load raw data from multiple sources and process them. This script is implemented using NodeJS along with various libraries such as xlsx, sharp, and yup. XLSX library was used to read the excel sheets and load the data into JSON format to process those data. Sharp library was used to load images, resize and process them according to the company's needs. YUP is a validation library to validate the data using the database schemas implemented by the company.

i Screenshots



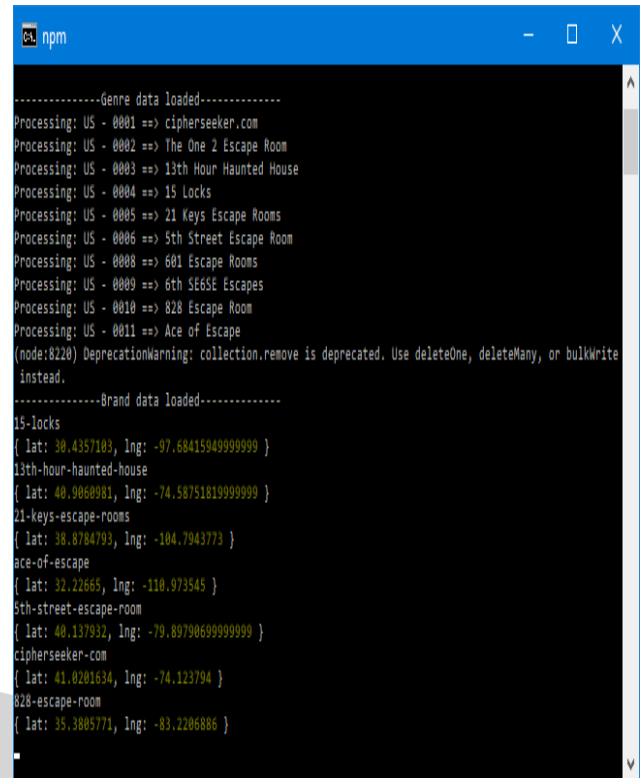
```

C:\WINDOWS\system32\cmd.exe - mongod
Microsoft Windows [Version 10.0.17134.885]
(c) 2018 Microsoft Corporation. All rights reserved.

C:\Users\pmani_000>mongod
2019-08-12T05:06:48.496-0700 I CONTROL [main] Automatically disabling TLS 1.0, to force-enable
e TLS 1.0 specify --sslDisabledProtocols 'none'
2019-08-12T05:06:48.678-0700 I CONTROL [initandlisten] MongoDB starting : pid=13620 port=2701
7 dbpath=C:\data\db\ 64-bit host=DesiGal-PC
2019-08-12T05:06:48.679-0700 I CONTROL [initandlisten] targetMinOS: Windows 7/Windows Server
2008 R2
2019-08-12T05:06:48.680-0700 I CONTROL [initandlisten] db version v4.0.6
2019-08-12T05:06:48.681-0700 I CONTROL [initandlisten] git version: caa42a1f75a56c7643d0b68d3
880444375ec42e3
2019-08-12T05:06:48.681-0700 I CONTROL [initandlisten] allocator: tcmalloc
2019-08-12T05:06:48.681-0700 I CONTROL [initandlisten] modules: none
2019-08-12T05:06:48.681-0700 I CONTROL [initandlisten] build environment:
2019-08-12T05:06:48.682-0700 I CONTROL [initandlisten] distmod: 2008plus-ssl
2019-08-12T05:06:48.682-0700 I CONTROL [initandlisten] distarch: x86_64
2019-08-12T05:06:48.682-0700 I CONTROL [initandlisten] target_arch: x86_64
2019-08-12T05:06:48.682-0700 I CONTROL [initandlisten] options: {}
2019-08-12T05:06:48.814-0700 I STORAGE [initandlisten] Detected data files in C:\data\db\ cre
ated by the 'wiredtiger' storage engine, so setting the active storage engine to 'wiredtiger'.
2019-08-12T05:06:48.815-0700 I STORAGE [initandlisten] wiredtiger_open config: create,cache_s
ize=1501M,session_max=20000,eviction=(threads_min=4,threads_max=4),config_base=false,static

```

Fig. 7 Screenshot of MongoDB Server Initializing



```

npm
-----Genre data loaded-----
Processing: US - 0001 ==> cipherseeker.com
Processing: US - 0002 ==> The One 2 Escape Room
Processing: US - 0003 ==> 13th Hour Haunted House
Processing: US - 0004 ==> 15 Locks
Processing: US - 0005 ==> 21 Keys Escape Rooms
Processing: US - 0006 ==> 5th Street Escape Room
Processing: US - 0008 ==> 601 Escape Rooms
Processing: US - 0009 ==> 6th SE6SE Escapes
Processing: US - 0010 ==> 828 Escape Room
Processing: US - 0011 ==> Ace of Escape
(node:8220) DeprecationWarning: collection.remove is deprecated. Use deleteOne, deleteMany, or bulkWrite
instead.
-----Brand data loaded-----
15-locks
{ lat: 30.4357103, lng: -97.68415949999999 }
13th-hour-haunted-house
{ lat: 40.9060981, lng: -74.58751819999999 }
21-keys-escape-rooms
{ lat: 38.8784793, lng: -104.7943773 }
ace-of-escape
{ lat: 32.22665, lng: -110.973545 }
5th-street-escape-room
{ lat: 40.137932, lng: -79.89790699999999 }
cipherseeker-com
{ lat: 41.0201634, lng: -74.123794 }
828-escape-room
{ lat: 35.3805771, lng: -83.2206886 }

```

Fig. 8 Screenshot of Data Processing Script Running in Command Prompt

IV. AUTOMATION TOOLS

A. Implementation and results

The scripts mentioned in this chapter have been implemented using PYQT5 library for Graphical User Interface (GUI) and Selenium for automation.

i. Main UI

This script is a basic script which just displays all the tools available to use by the user. It is built over PYQT5 library and each button is configured to open the specific tool.

ii. Improvements to proposed code

The python development team in the company had developed a series of tools for automation purposes. But these scripts posed a lot of bugs and were to be fixed and the performance was improved by using threading concepts.

iii. Converting Python to EXE

The tools created were too complex to use as keeping track of each program was very hard for end user. Thus, all the tools were converted into a single EXE file for easier usage and portability of the software built. A library called CX_Freeze was used to assist the process of converting Python scripts into executable file.

Results

i. Screenshots

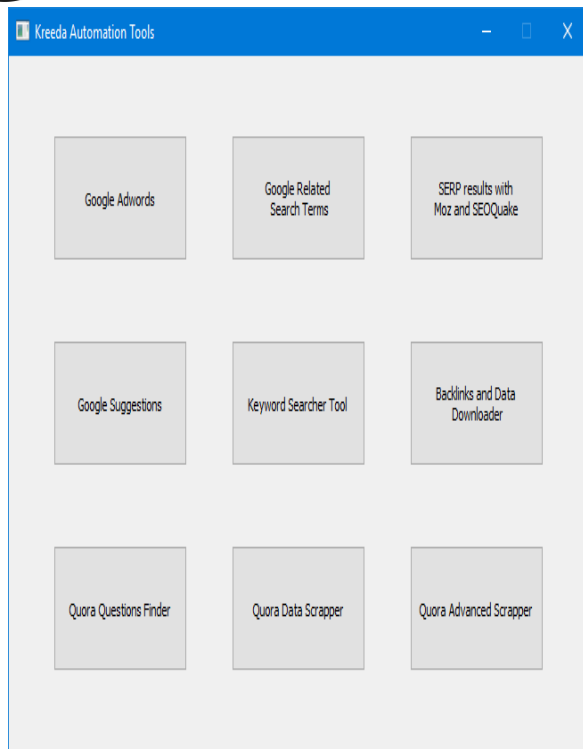


Fig.9 Screenshot of Main UI

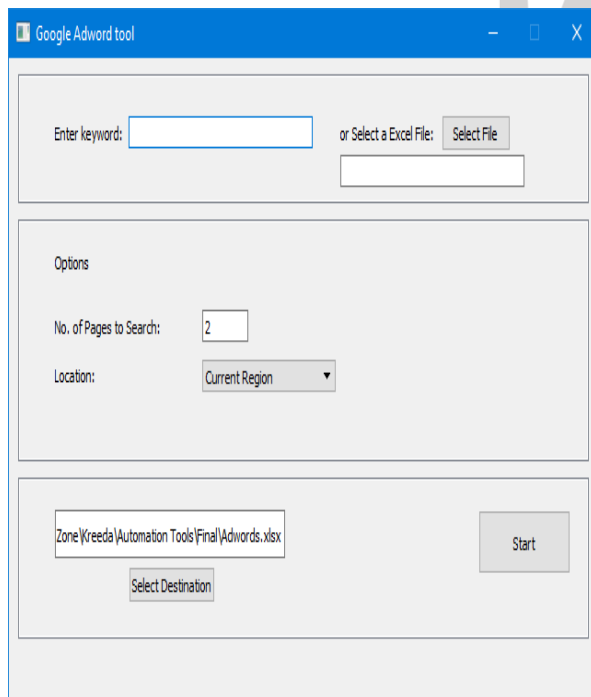


Fig.10 Screenshot of Adwords Mining Tool

V. CONCLUSION AND FUTURE WORK

A. Conclusion

In the public cloud industry, the sector has grown enormously. In contrast to private clouds that organizations use externally, almost anyone with a credit card can access public cloud technologies. Following this pattern, companies are increasingly implementing public cloud computing services such as Amazon Web Services

(AWS), Google Cloud Platform (GCP), Microsoft Azure, IBM Bluemix, Alibaba cloud, and Oracle cloud. While the spectrum of public cloud providers promotes healthy competition, it is difficult to select the best cloud.

We recently conducted a comprehensive assessment of public cloud providers. This study results in a systemic evaluation that finest characterizes cloud effectiveness measures. In order to meet the application standards correctly, we can quickly pick the most effective provider for our execution. One other problem is the time-consuming method that plagues companies of all dimensions of the manual entry of the data in Excel sheets. The data accessed digitally may contribute to prevalent mistakes without enough accounting. These mistakes may seem insignificant but may cost your company thousands of bucks. We have written python scripts that detect such errors and report the errors made by data entry staff to remedy them promptly. Finally, we concluded that the proposed enterprise grade level cloud architecture will provide end users for deployment of cost-effective application and exploration of valid information related to specific domain.

B. Future Work

The raw data collected by the mining tools and scripts created can be used for further research work and analysis such as sentiment analysis. The domain of feeling assessment is an interesting fresh study direction because a great many apps in the real world are essential for stronger decision-making when it comes to finding people's opinions. One of the important parts of this field is the creation of methods for document-level sentiment analysis. People have recently begun to express views on the Internet that have enhanced the need to analyze the views contained online for several apps in the real world. A great deal of studies is available in literature to detect feelings from the text. Nevertheless, the current sentiment assessment models are being greatly improved.

REFERENCES

- [1] Ang Li, Xiaowei Yang, Kandula, S., & Ming Zhang. (2011). Comparing Public-Cloud Providers. IEEE Internet Computing, 15(2), 50–53.doi:10.1109/mic.2011.36
- [2] Peng, J., Zhang, X., Lei, Z., Zhang, B., Zhang, W., & Li, Q. (2009). Comparison of Several Cloud Computing Platforms. 2009 Second International Symposium on Information Science and Engineering.doi:10.1109/isise.2009.94
- [3] Gmach, D., Rolia, J., & Cherkasova, L. (2012). Comparing efficiency and costs of cloud computing models. 2012 IEEE Network Operations and Management Symposium.doi:10.1109/noms.2012.6211977

- [4] Heyong Wang•Wu He•Feng-Kwei Wang,"Enterprise cloud service architectures",
<https://www.researchgate.net/>
- [5] N. Ram Ganga Charan, S. Tirupati Rao and Dr .P.V.S Srinivas,"Deploying an Application on the Cloud",in (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 5, 2011.
- [6] Asadallah Najafi , Islamic Azad University, Zanjan Branch, Moallem Ave., Zanjan, Iran, Model Validation and Verification of Data Mining from the Knowledge Workers Productivity Approach.
- [7] Prakash Gopalakrishnan, B. Uma Maheswari , "Research On Enterprise Public and Private Cloud Service" in International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8, Issue- 6S4, April 2019
- [8] JOSE MOURA and DAVID HUTCHISON,"Review and Analysis of Networking Challenges in Cloud Computing"
- [9] Amazon Web Services, <https://aws.amazon.com/what-is-aws/>
- [10] Rabi Prasad Padhy,Manas Ranjan Patra,"Architecture & Design of Affordable and Highly Available Enterprise Cloud Service", International Journal of Cloud Computing and Services Science (IJ-CLOSER),Vol.2, No.2, April 2013, pp. 85~105,ISSN: 2089-3337.
- [11]"Aaqib,Amit Chaturveti", Rashid,"Cloud Computing Characteristics and Services: A brief Review",International Journal of Computer Sciences and Engineering.

