**School of Information Technology and Engineering (SITE)**

B.Tech (Information Technology)

Course Project Report

# AccountantPro:CA platform to manage transcripts and delegate employees to clients

Submitted for the Course:
ITE 3007 : Cloud Computing and Virtualization

Offered by Dr. R. K. NADESH during SUMMER SEMESTER
TERM 5 - 2022

***By:***
**Aayu Ojha 19BIT0225**
**Dipesh Balani 19BIT0354**
**Utkarsh Goyal 19BIT0402**

AUGUST 2022

School of Information Technology & Engineering B.Tech (InformationTechnology)

SUMMER SEMESTER TERM 5 - 2022
ITE 3007 : Cloud Computing and Virtualization

**A Report on the Course Project**
**AccountantPro:CA platform to manage transcripts and delegate employees to clients**

TEAM Name : **DEADPOOL**

Team Members:

1)Dipesh Balani; 19BIT0354; 9377390528; dipesh.vijaykumar2019@vitstudent.ac.in
2)Utkarsh Goyal; 19BIT0402;7742027100;utkarsh.goyal2019@vitstudent.ac.in
3)Aayu Ojha; 19BIT0225; 7280891010; aayu.ojha2019@vitstudent.ac.in

Project Title : AccountantPro:CA platform to manage transcripts and delegate employees to clients

### 1.1 Abstract and Background

Managing the finances of their customers and suppliers plays a significant role in the work of chartered accounting businesses. They typically have to do this in order to keep track of all of their customers' business and personal invoices and costs.conduct evaluations, carry out the necessary assurance and audit, and financial accounting and reporting, applied finance, managerial accounting, or taxation.

### 1.2 Problem Statement

Much of the work of an accountant revolves around financial management of clients and suppliers. This typically requires tracking all client's professional / personal invoices and expenses as needed. This allows you to perform valuations, perform required audits and audits, accounting and reporting of

Treasury, management accounting, and applicable treasury or taxation. The current problem with this scenario is that all invoices managed by these companies are generally physically managed or have an internal system to handle them. There is no such open source platform in the community at your disposal. Costs or costs that a company can build according to its own specific requirements. The goal of our team is to bring AccountantPro to such platforms.

## 1.3 Related Works

### 2.1 Literature Survey (Should be elaborately discussed with its citation)

**1. https://www.sciencedirect.com/science/article/pii/S1319157818303999**

**BDNA-A DNA inspired symmetric key cryptographic technique to secure cloud computing**
**Year - 2018**

The paper proposes a new encryption technique for data on the cloud called BDNA. There are several algorithms for encryption of data and files on the cloud each with its advantages and disadvantages. In recent years the DNA encryption technique has become very popular as it provides a very high level of cloud storage security. But DNA Cryptography has many computational drawbacks, as high technology laboratories are required for the actual implementation of this technology. The automation of DNA processes cannot be done. For DNA synthesis, human reprocessing is required at every phase which has become a stumbling block in its extensive development.
The proposed algorithm of BDNA stands for Binary DNA cryptography and it is a pseudo DNA cryptography algorithm. Unlike DNA it does not require high computational resources but also stands out to be better than the other cryptography algorithms in many aspects.
The security framework on receiving the file encrypts it using the newly proposed symmetric key encryption technique (BDNA).The framework asks the data owner for the secret key for encrypting the data. Then the file is uploaded onto the cloud for storage. Further, the framework encrypted this secret key with the public key of all the authorized users provided by the data owners. The secret keys are encrypted using the public key of the authorized users. The key management takes place at the security framework module. Whenever a user needs to access the file, he/she enters the authentication credentials at the framework's interface. The security framework retrieves the corresponding file from the cloud and the user is then prompted to enter his private key using which the secret key is decrypted. Using this secret key, the file is encrypted by DNA decryption technique and is sent to the user.
The encryption module uses two values one the user's secret and another a random number, so on each iteration the value changes unpredictably. The proposed algorithm is secure against chosen plaintext attacks (CPA). A symmetric algorithm is said to be secure against CPA if, for all the probabilistic polynomial-time, the adversaries have

negligible chances of guessing the right plaintext.
Talking about performance, in comparison to AES, DES, BF and DNA cryptography algorithms, BDNA has the least ciphertext size, highest throughput and least encryption time in seconds.

**2. https://ieeexplore.ieee.org/document/8794794**

**Energy-Efficient Load Balancing Ant Based Routing Algorithm**
**Year - 2019**

The paper proposes Energy-Efficient Load Balancing Ant-based Routing Algorithm (EBAR) for Wireless Sensor Networks. EBAR adopts a pseudo-random route discovery algorithm and an improved pheromone trail update scheme to balance the energy consumption of the sensor nodes. It uses an efficient heuristic update algorithm based on a greedy expected energy cost metric to optimize the route establishment. Finally, in order to reduce the energy consumption caused by the control overhead, EBAR utilizes an energy-based opportunistic broadcast scheme.
WSNs are desired in the paper as undirected weighted graphs. Initially a heuristic update procedure is launched to find the path increasing the expected energy cost greedily. At intervals a forward ant is launched from source to sink node. The ant generates the route by using a probabilistic state transition rule and once it reaches the sink node, it backtracks the route.In the ACO framework, heuristic values represent a priori information about the problem or run-time information provided by a source different from the ants . The heuristic information is usually defined as a function of the residual energy or the average energy of neighbor nodes in ant-based routing. Making routing decisions on these functions aims to increase the energy efficiency of the WSNs.
In comparison with other methods such as IACO, EEABR and Sensorant, EBAr was found to have higher throughput, lower average energy consumption and lower packet overhead in the majority of cases for both static and dynamic networks. However, in In very specific cases of dynamic network IACo did outperform in terms of throughput.

**3. https://ieeexplore.ieee.org/abstract/document/8852632**

**Load Balancing and Server Consolidation in Cloud Computing Environments**
**Year - 2019**
This paper surveys published load balancing algorithms that are achieved by server consolidation via a meta-analysis. Load balancing with server consolidation enriches the exploitation of resource utilization and can enhance Quality of Service (QoS) metrics, since data-centers and their applications are increasing exponentially. This meta-study reviews the literature on load balancing and server consolidation and presents a ready reference taxonomy on the most efficient algorithms that achieve load balancing and server consolidation.
The paper identifies the challenges of load balancing and server consolidation in cloud computing as VM migration, cloud node's geographical distribution, emergence of small data centers and energy consumption. It keeps the metrics for evaluation as response time, performance, makespan, throughput, resource utilization, migration

time, scalability, degree of imbalance, fault tolerance, energy consumption and carbon emission.
It compares the three major methods used for server consolidation or load balancing namely, the Exact method, the Heuristics Method and the Meta Heuristics Method over the performance metrics mentioned above and then compares their evolution over the years based on these metrics.
It also suggests some future research directions regarding how to estimate the accuracy of incoming workloads, adaptive server thresholds instead of keeping them fixed, security aware migration threats, some other nature-inspired techniques and dwindling resource utilization.

**4. https://ieeexplore.ieee.org/document/8683979**

**Osmotic Bio-Inspired Load Balancing Algorithm in Cloud Computing**
**Year - 2019**

The goal of this study is to present a hybrid metaheuristics technique that incorporates osmotic behavior and bio-inspired load balancing algorithms. The osmotic behavior allows for the automatic deployment of virtual machines (VMs) across cloud infrastructures. The paper uses the advantages of these bio-inspired algorithms to build an osmotic hybrid artificial bee and ant colony (OH BAC) optimization load balancing method after the hybrid artificial bee colony and ant colony optimization established its efficiency in a dynamic setting in cloud computing.
It addresses the shortcomings of existing bio-inspired algorithms for load balancing amongst physical machines. When compared to existing methods, the simulation results show that OH BAC reduces energy usage, the number of VM migrations, and the number of shutdown hosts. Furthermore, it improves service quality (QoS), which is assessed by service level agreement violations (SLAV) and performance degradation due to migrations (PDMs).
OH_BAC inherits ACO's core behaviors of swiftly discovering solutions at diversity systems and ABC's behavior of waggle dancing to share information. Waggle dance is represented as a knowledge base in OH BAC. Instead of selecting PMs at random, OH BAC uses a knowledge base and the osmosis technique to filter them according to energy usage. The dynamic value of the threshold is taken into account by OH BAC depending on the status of the cloud system.
OH_BAC achieves better results than H_BAC and other algorithms. This is due to OH_BAC activates the most suitable osmotic host among all PMs in the system to decrease power consumption.OH_BAC gets more enhancement than H_BAC and ABC. This is due to ACO and ABC together in OH_BAC select the most suitable VM to migrate from the most suitable overloaded host.OH_BAC improves energy consumption, SLAV, number of VM migrations, and number of hosts' shutdowns with compared to other algorithms. However, it has more SLATAH with compared to other algorithm but it is not affected in the performance of the cloud system. OH_BAC achieves improvements by about 27% compared with H_BAC, 21% compared with ABC algorithm, and 18% compared with ACO algorithm. It is shown that OH_BAC also has a minimum value of processing time

**5.https://ieeexplore.ieee.org/document/8344738**

**Comparative Study of Different Cryptographic Algorithms for Data Security in Cloud Computing**
**Year - 2017**

This study compares various cryptographic encryption algorithms in terms of their various key features, and then evaluates their performance costs based on a set of key requirements. DES, 3DES, IDEA, CAST128, AES, Blowfish,RSA,ABE, and ECC are some of the algorithms used.
The paper provides a description of the most common types of cryptographic algorithms used in cloud storage cryptography. There are two major categories of cryptographic algorithms: Symmetric Key Algorithms and Asymmetric Key Algorithms. It then discusses the major types of Symmetric key algorithms and their workings in brief, namely, Data Encryption Standard (DES) algorithm, Triple Data Encryption Standard Algorithm, Carlisle Adam &Stafford Tavares [CAST-128] algorithm, International Encryption Algorithm, Advanced Encryption Standard Algorithm and Blowfish Encryption Algorithm. Then it discusses the major types of Asymmetric key algorithms and their workings in brief, namely, Rivest Shamir Adlemen [RSA] algorithm, Attribute Based Encryption algorithm, Elliptic Curve Cryptography algorithm and Homomorphic Encryption Algorithm.
It then compares all of the above mentioned algorithms based on their security levels and the attacks they are vulnerable to. Out of all these AEs, Blowfish and RSA algorithms are found to have the best security levels and are prone to least severe and types of attacks. The paper then compares the DES,3DES,AES,Blowfish & RSA algorithms based on the evaluation metrics of Encryption Time,Decryption time, Avalanche Effect and Memory consumption
RSA is found to have the lowest encryption timea and 3DES the highest. Initially AES was consuming more encryption time as compared to other encryption algorithms when size of input file was less but when we try to implement with large size input file, RSA proves to take highest encryption time & Blowfish takes least time for the same input.RSA proves to take the highest time for decryption, and blowfish takes least time for the same input. Further the implementation reveals that RSA consumes maximum amount of memory & Blowfish consumes least with respect to unit operation.
It was concluded that Blowfish is best in terms of memory requirement, whereas RSA has a large memory requirement, so blowfish can fit well in small applications specially in embedded applications & for devices with small memory. The avalanche effect of AES is high,so AES can be preferred for application where privacy and integrity of the message is of top priority.

**6. https://ieeexplore.ieee.org/document/8815947**

**Cryptographic Solutions for Cloud Storage: Challenges and Research Opportunities**
**Year - 2019**

This paper provides a critique of the wide range of cryptographic schemes designed for securing sensitive data in the cloud computing environment, as well as outlining the research opportunities in the use of cryptographic techniques in cloud computing.

The paper first discusses the various approaches available to ensure the reliability of data stored on the cloud. The various techniques it discusses are Proofs of Storage, Proofs of Location and Redundancy Proofs of Storage.
Next it discusses the various techniques involved in secure data sharing. The ones the paper discusses are Fine GRained Access Control, Secure Data Sharing Delegation and Keyword Search.
It then goes on to discuss the models of all the above mentioned techniques in detail and explains their working. PoS system model, which consists of data owners, CSP, and an optional third party auditor (TPA). Data owners upload their data to the CSP via a secure data flow, and the data owner or the TPA can audit the integrity of the data. A PoL protocol is usually a challenge-response protocol that runs between a verifier (e.g., a data owner or a TPA) and a CSP that manages several cloud storage servers. It also compares the various techniques used in PoL and PoS against their AC2 scores. In a redundancy PoS system (see Fig. 4), a user (data owner) stores the data with a CSP that maintains several storage servers distributed at different locations. The CSP has to retain at least one copy of the user's data in each agreed geolocation.
It then goes on to discuss the various data sharing techniques mentioned above. Most of these are based on ABE (Attribute based encryption) techniques.

**7. https://ieeexplore.ieee.org/abstract/document/8219724**

**Auditing Big Data Storage in Cloud Computing Using Divide and Conquer Tables Year - 2017**

This study provides an efficient RDC approach based on algebraic features of outsourced files that incur the least amount of compute and communication cost. This paper's main contribution is the introduction of a new data structure known as the Divide and Conquer Table (D&CT), which effectively accommodates dynamic data for files of normal size. Furthermore, because of this data structure, our technology can be used for large-scale data storage with low processing costs. The suggested technique and existing methods have considerable differences in terms of calculation and communication cost on the auditor and cloud, according to the one-way analysis of variance.
The paper presents an efficient way for remotely verifying the integrity of data in cloud storage using the outsourced files' algebraic signatures. On the client and cloud sides, our data auditing solution incurs the least computation and communication costs. It created D&CT as a new data structure to handle dynamic data operations like insert, add, delete, and edit with ease. D&CT enables our RDC scheme to work with a wide range of file sizes (normal and large-scale files) while requiring the least amount of processing on the client and cloud sides. It has created a scheme to demonstrate the security of our technology, justify its performance for typical and large-scale files, and compare it to current data auditing methods. We also use one-way analysis of variance (ANOVA) to validate the comparison and show that there are substantial differences between the proposed method and the existing methods.
In comparison of evaluation metrics such as computation time and communication cost, with the methods developed by Wand and Yang, the scheme suggested by this paper either proves better than both or is at par with them in variou operations. The only place where Wangs scheme outperforms this is verification on Server side.

**8.** https://link.springer.com/article/10.1007/s12083-020-00959-6

**FAST: Fast Accessing Scheme for data Transmission in cloud computing**
**Year - 2020**

To address the issues regarding data transmission on cloud, a novel access control mechanism is provided in this study. For rapid and efficient data access, the Cloud Service Provider (CSP) creates a temporary table depending on the data type and popularity value of the DO. By using the table, the cloud service provider can quickly find the data owner, reducing data access time significantly. The suggested scheme's experimental findings and theoretical analysis demonstrate its superiority to existing schemes.
In the proposed FAST, the CSP maintains a table based on the data type of the DO and the popularity value of the DO for fast and efficient data accessing. 2 The proposed scheme minimizes the searching time of PCKDO, data accessing time and CPU utilization. Moreover, the proposed ACM also reduces the system overhead. 3 Theoretical analysis and experimental results of FAST have been presented for proving its efficiency.
The proposed system model of FAST is composed of three entities namely, the Cloud Service Provider, the Data owner and the User. In the proposed FAST, the CSP maintains a temporary CSP Table (CSP-TAB). In the CSP-TAB, there are six attributes: (i) Group Identity (ID) (GP_ID) (ii) Data Type (DT) (iii) Popularity Value (PV) (iv) Data Owner's ID (DO_ID) (v) DO's Time & Date (DO_T&D), and (vi) Group's Time & Date (GP_T&D). The GP ID keeps track of all of the groups' identities. The data types of the DOs are kept by DT. Each DO's popularity value is maintained by PV. The PV of a DO increases by one when it sends data to a user. The PV of any DO can be between 1 and 10 in FAST, i.e. the PV of any DO can be between 1 and 10. The CSP chooses the data type and threshold PV of the DO based on its convenience. There may be several DOs that share the same data. As a result, they are stored in the same DO ID. When the DO has shared the requested data, the DO T&D is updated with the appropriate date and time. The group formation details are stored in GP T&D. The CSP adds the DO to the table based on the data type whenever DO shares data on the cloud server. The CSP searches the DO from the CSP-TAB based on the user's requested data type and PV at the moment of data access. If the DO already exists in the table and the PV of the particular DO has not yet reached the threshold value, the CSP increases the PV of the DO by one. Otherwise, a new GP ID with the DO's details is created. The CSP then hands the PCKDO over to the user. As a result, the DO with the threshold PV of a specific data type is placed at the top of the DO ID, and the CSP does not need to search the entire database for the PCKDO. The user can send a request in less time to the DO in order to get the credentials by using the PCKDO. Thus, the CSP-TAB can help to reduce the searching time of a DO as well as the data access time.
The worst-case complexity of insertion algorithm, deletion algorithm and searching of key algorithms is O(PQ).

**9.** https://ieeexplore.ieee.org/document/8693725

**Energy Efficient Virtual Machine Placement with An Improved Ant Colony Optimization over Data Center Networks**
**Year - 2019**
The paper suggests an energy-efficient virtual machine placement technique in this paper, with the goal of lowering communication costs and power consumption in traffic-aware data center networks. An enhanced ant colony optimization with adaptive parameter setting was presented to balance its fast convergence and robust search capabilities in order to address such an optimization challenge. Simulation findings showed that our method outperformed previous algorithms in terms of power usage and communication cost, as well as a significant reduction in run time under various traffic patterns and settings.
The VMP problem has been demonstrated to be an NP-hard problem, meaning that Finding a solution in polynomial time is difficult. As a result, obtaining answers necessitates the use of efficient and desired algorithms. An improved ant colony optimization with adaptive parameter setting, dubbed APACO, is suggested to improve the convergence rate and search capabilities of our energy efficient VMP scheme (E2VMP) with bi-objective function.
The ants make use of the heuristic information to select their path ahead which makes them the intelligent ant. This algorithm hence doesn't follow the greedy algorithm and hence is more energy efficient. Other ants follow what is called the phenomenal trail. A phenomenal diffusion model is applied to simulate a real ant colony. Considering the diffusion, we can improve the collaboration among ants and reduce the times of iteration.The Pareto approach is more often used to cope with multi-objective problems than the conventional weighted formula approach. The solution set is known as a non-dominated solution set.
The simulation results show that the power consumption and communication cost are increasing when the number of VMs is larger. In the same network topology, due to the increase of VMs, more PMs are in the active model, which will cost more energy. Also, when more VMs work, the resource requirements will increase, leading to a large communication cost. The ACO algorithm outperforms the FFD algorithm, because the ACO algorithm can search the solution space more efficiently and globally. Also the time decreases with the increase of bandwidth.It was also seen that the metaheuristic optimization algorithms (ACO and AP-ACO) spend a longer time than those of FFD algorithm and Random.

**10.** **https://www.sciencedirect.com/science/article/abs/pii/S0045790616300544**

**Cloud security: Emerging threats and current solutions**
**Year - 2016**

This article includes a comprehensive analysis of the primary threats preventing widespread adoption of cloud computing, as well as a straight-to-the-point overview of the current solutions offered by the leading suppliers. By offering a picture of the main research trends and most accredited methodologies, the study also presents the (near) future directions of cloud security research. The research is based on a best-of-breed mix of proprietary and open-source cloud solutions. The report can thus be used by IT employees to obtain a better understanding of the security issues associated with cloud computing and to rapidly balance the benefits and drawbacks of current state of

the art solutions.
The paper states the major security threats nowadays are Shared technological vulnerabilities, Data breaches, Account or service traffic hijacking, DoS (Denial of Service) attacks and Malicious insider attacks. It states that the aforementioned open issues can be caused by three main vectors of attack: Network, Hypervisor, and Computing Hardware. Three types of attackers map on these vectors: external users, internal users, and the cloud provider itself (embodied in a malicious employee). It then discusses the various techniques such as filtrations and other countermeasures such as IDPS, IDs, HIDS, NIDS and DIDS for as countermeasures to keep check on system availability. To prevent data confidentiality and integrity it then discusses the various protocols of SSL and TLs over which data should be communicated and other data encryption techniques to be used. It then talks about the various commercial solutions provided by CSPS like AWS, Azure, GCP etc to counter these attacks.
It then also discusses countermeasures against hardware based attacks such as device isolation, network isolation etc. and the commercial counterparts adopted by AWs, Azure, DO, GCP etc.
According to the analysis in the paper, Amazon AWS and VMWare vCloud both offer a range of products and mechanisms that are provided either directly by the CPs or through third-parties. It turns out that, although certain security solutions have attained a certain level of technical maturity, others are still in their infancy and are not suited for deployment in a production environment. There are numerous approaches for dealing with IS1 and IS2, however they all need to be improved. The most advanced countermeasures against network-based assaults are those at the network level. Instead, application-level techniques to handle IS3 and IS4 are still being refined, and many researchers are concentrating on them.

**11.** **https://ieeexplore.ieee.org/document/8763773/authors#authors**

**SSH Key Management Challenges and Requirements**
**Year - 2019**

This paper highlights the development in SSH key management, the problems SSH is able to solve and It also discusses further solutions using SSH keys that can help us in solving more critical infrastructure problems related to internet applications.
Currently, the most popular use case of SSH keys is user authentication, by detecting existing keys by scanning the installed system, verifying their actual users, and ensuring proper termination and provisioning of user access through those keys.
Though for enterprise systems, using just SSH keys can become a risky approach as any vulnerability in any of the SSH keys and cause a breach in multiple systems that uses that particular SSH key, so an Identity and Access Management System is used for the provision of user access.
There is no automated system for deploying a host while it's installing to an SSH key management system. Many clients yet do not support OpenSSH Certificates and hence yet to be standardized. These are some of the issues which need to be solved with SSH Key management.

**12.** **https://ieeexplore.ieee.org/abstract/document/9213745**

**Cloud Cryptography: User End Encryption**
**Year - 2020**

This paper discusses the vulnerability and integrity of user data present in the cloud. It explains solutions using user-end cryptography of data to protect users from breaches in the cloud data centers. Symmetric key cryptography is the algorithm proposed by the paper for implementation, as its the users who are present at both the ends of encryption and decryption and it also eliminates the need of key sharing for the users. The key is generated using the binary value of ciphertext obtained from the encryption phase. The decryption procedure involves extraction of ciphertext and then converting to ASCII code character, and then to a binary value.
The solution provided by the paper is limited to text-based content. For audio, video, and other types of content more advanced techniques of encryption is required which is not covered in the paper.

**13. https://ieeexplore.ieee.org/document/8446915**

**Virtual Machine Migration Algorithm for Cloud Data Center**
**Year - 2017**

This paper discusses migration techniques for applications to be migrated to virtual machines in the Cloud Data Centers. It proposes two kinds of migration models for this purpose. One is a dynamic double threshold decision model which uses allocated resources based on when the current host is active. The other model is the destination choice model which takes into consideration the availability of cloud data services and uses the energy efficiently by optimized usage of the cloud data services, by using the most relevant virtual machine.
Comparison of performance between these models and the usual migration algorithm shows that the one proposed by the paper greatly minimizes the energy consumption.

**14.https://ieeexplore.ieee.org/document/7416240**

**Multiagent-Based Resource Allocation for Energy Minimization in Cloud Computing Systems**
**Year - 2017**

This paper discusses the problem of allocating virtual machines to software applications in an efficient manner in order to reduce energy consumption. It introduces a Decentralized Multiagent that allocates virtual machines in a decentralized manner, unlike other standard methods which allocate in a centralized way without considering migration costs.
The proposed solution works by assisting the PM in handing virtual machines through a cooperative agent. These agents allocate the VMs by using an auction-based mechanism and also minimizes migration costs of Virtual machines using VM consolidation mechanisms.
Analysis showed that this system helps in efficient energy consumption more than other mechanisms.

**15.** https://ieeexplore.ieee.org/document/8386503

**Research and design of cryptography cloud framework**
**Year - 2018**

This paper introduces a cryptography framework that can be used for cryptographic services in a cloud computing environment. The framework is built from two aspects, and those are the function of service and module flow which improves the flexibility of cryptography in a cloud environment. It is made out of a cluster of cloud cipher machines to make a centralized computing core. The platform used for managing the cryptographic mechanisms are made of VCM management scheduler, a cloud cipher machine monitor, network manager, and authentication module
Through the method proposed in the paper, a very high level of availability of the a cryptographic mechanism was achieved.

**16.** https://ieeexplore.ieee.org/document/8121690

**Improved Data Storage Confidentiality in Cloud Computing Using Identity-Based Cryptography**
**Year - 2017**
This paper proposes a way to securely store data in cloud data centers based on Identity Based Cryptography. It also allows authenticated users to access their data securely. The architecture of the proposed system is composed of Cloud Storage Providers, Users, Private Key generators. The Private key generator generates all, the parameters needed for the key generation, and then using the unique ID of every user creates a unique private key for every user.
Any new user can obtain their key by sending their unique id to the system. Whenever a user stores data on the cloud, his unique private key is used to encrypt all his data before it is stored on the cloud, This ensures that even the cloud service provider is not capable of accessing its user's data.
If the user wants to share access details if a particular file, a public key for the file can be generated using the private key of the user, which then can be shared with any person who can access the file using the public key.

**17.** https://ieeexplore.ieee.org/document/9202055

**Research and Application of Data Privacy Protection Technology in Cloud Computing Environment Based on Attribute Encryption**
**Year - 2020**

This paper proposes an encryption mechanism based on the weight of attributes and attributes authority to enhance the security systems of cloud data centers. It uses a proxy encryption algorithm to generate the key, which takes a few parameters as input. These parameters include the user's authority weight and other access details which determine the strength of the key generated by the algorithm. The algorithm also takes as input the master key and the property set RMA. DekProxyKeyGen algorithm is used to create symmetric key's proxy re-encryption keys which are used to re-encrypt any decrypted data.

This solution is applicable for systems that do not have complex policy requirements such as biometric systems.

**18.** https://ieeexplore.ieee.org/document/8007315

**Latent-Data Privacy-Preserving With Customized Data Utility for Social Network Data**
**Year - 2018**

This paper discusses the privacy concerns of big data technologies specifically in a social network setup and proposes a data sanitization strategy that can be also used in a cloud computing environment in order to get the best benefits out of collected data while also preserving the privacy rights of users.
It suggests sanitizing data using either Attribute sanitization method or Link Sanitization Method. In Attribute sanitization, data is sanitized by either adding attributes, removing attributes, or replacing an attribute with another.
In the Link sanitization method, data is sanitized by either adding new links or removing existing links. Then the data is optimized for maximum latent data piracy.

**19.** https://ieeexplore.ieee.org/document/8006268

**Predicted Affinity Based Virtual Machine Placement in Cloud Computing Environments**
**Year - 2020**

This paper discusses the relationship between any two virtual machines based on the requirements of resources given by ARIMA prediction. It also explains an affinity model which calculates the volatility of resource utilization after putting any two virtual machines on the same host. This model is called the Predicted Affinity-based Virtual Machine Placement Algorithm, which is used to place any virtual machines that have an affinity relationship on the same physical machine.
An experiment was also run to compare other VM placement algorithms with the one discussed in the paper which showed the latter one as the better performer, though for complex systems the prediction algorithm needs to be further developed.

**20.** https://ieeexplore.ieee.org/document/6149560

**Cloud Computing Security: From Single to Multi-clouds**
**Year - 2012**

This paper discusses the possible security threats in a multi-cloud environment and steps needed to be taken for protection against the same. The various risks involved in multi-cloud which are discussed are configuration errors, application hardening, data governance, data visibility, shared security model.
The paper suggests hash function as one of the solutions to data integrity. In cases of large amounts of data, a hash tree function can be used such as SIRUS and TDB. It also suggested Proof of Retrievability and Proof of Data Possession protocols to be implemented for ensuring that users can always retrieve their old data in case of a data

breach in the cloud data centers.

**REFERENCES**


1. Manreet Sohal, Sandeep Sharma,BDNA-A DNA inspired symmetric key cryptographic technique to secure cloud computing,Journal of King Saud University - Computer and Information Sciences,2018

2. X. Li, B. Keegan, F. Mtenzi, T. Weise and M. Tan, "Energy-Efficient Load Balancing Ant Based Routing Algorithm for Wireless Sensor Networks," in IEEE Access, vol. 7, pp. 113182-113196, 2019, doi: 10.1109/ACCESS.2019.2934889.

3. M. Ala'Anzy and M. Othman, "Load Balancing and Server Consolidation in Cloud Computing Environments: A Meta-Study," in IEEE Access, vol. 7, pp.141868-141887, 2019, doi: 10.1109/ACCESS.2019.2944420.

4. M. Gamal, R. Rizk, H. Mahdi and B. E. Elnaghi, "Osmotic Bio-Inspired Load Balancing Algorithm in Cloud Computing," in IEEE Access, vol. 7, pp. 42735-42744, 2019, doi: 10.1109/ACCESS.2019.2907615.

5. P. Semwal and M. K. Sharma, "Comparative study of different cryptographic algorithms for data security in cloud computing," 2017 3rd International Conference on Advances in Computing,Communication & Automation (ICACCA) (Fall), 2017, pp. 1-7, doi: 10.1109/ICACCAF.2017.8344738.

6. L. Zhang, H. Xiong, Q. Huang, J. Li, K. R. Choo and J. LI, "Cryptographic Solutions for Cloud Storage: Challenges and Research Opportunities," in IEEE Transactions on Services Computing, doi: 10.1109/TSC.2019.2937764.

7. M. Sookhak, F. R. Yu and A. Y. Zomaya, "Auditing Big Data Storage in Cloud Computing Using Divide and Conquer Tables," in IEEE Transactions on Parallel and Distributed Systems, vol. 29, no. 5, pp. 999-1012, 1 May 2018, doi: 10.1109/TPDS.2017.2784423.

8. Namasudra, S., Chakraborty, R., Kadry, S. et al. FAST: Fast Accessing Scheme for data Transmission in cloud computing. Peer-to-Peer Netw. Appl. 14, 2430–2442 (2021).

9. W. Wei, H. Gu, W. Lu, T. Zhou and X. Liu, "Energy Efficient Virtual Machine Placement With an Improved Ant Colony Optimization Over Data Center Networks," in IEEE Access, vol. 7, pp. 60617-60625, 2019, doi: 10.1109/ACCESS.2019.2911914.

10. Luigi Coppolino, Salvatore D'Antonio, Giovanni Mazzeo, Luigi Romano,Cloud security: Emerging threats and current solutions,Computers & Electrical Engineering,Volume 59,2017,Pages 126-140

11. T. Ylonen, "SSH Key Management Challenges and Requirements," 2019 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS), 2019,

pp. 1-5, doi: 10.1109/NTMS.2019.8763773.

12. S. A. Nooh, "Cloud Cryptography: User End Encryption," 2020 International Conference on Computing and Information Technology (ICCIT-1441), 2020, pp. 1-4, doi: 10.1109/ICCIT-144147971.2020.9213745.

13. Y. Liu, J. Gao and Y. Yao, "Research on Virtual Machine Migration Algorithm for Cloud Data Center," 2017 International Conference on Computer Systems, Electronics and Control (ICCSEC), 2017, pp. 1376-1381, doi: 10.1109/ICCSEC.2017.8446915.

14. W. Wang, Y. Jiang and W. Wu, "Multiagent-Based Resource Allocation for Energy Minimization in Cloud Computing Systems," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 47, no. 2, pp. 205-220, Feb. 2017, doi: 10.1109/TSMC.2016.2523910.

15. S. Lei, W. Zewu, Z. Kun, S. Ruichen and L. Shuai, "Research and design of cryptography cloud framework," 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), 2018, pp. 147-154, doi: 10.1109/ICCCBDA.2018.8386503.

16. H. Guesmi and L. A. Saïdane, "Improved Data Storage Confidentiality in Cloud Computing Using Identity-Based Cryptography," 2017 25th International Conference on Systems Engineering (ICSEng), 2017, pp. 324-330, doi: 10.1109/ICSEng.2017.32.

17. W. Zhang and S. Jin, "Research and Application of Data Privacy Protection Technology in Cloud Computing Environment Based on Attribute Encryption," 2020 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS), 2020, pp. 994-996, doi: 10.1109/ICPICS50287.2020.9202055.

18. Z. He, Z. Cai and J. Yu, "Latent-Data Privacy Preserving With Customized Data Utility for Social Network Data," in IEEE Transactions on Vehicular Technology, vol. 67, no. 1, pp. 665-673, Jan. 2018, doi: 10.1109/TVT.2017.2738018.

19. X. Fu and C. Zhou, "Predicted Affinity Based Virtual Machine Placement in Cloud Computing Environments," in IEEE Transactions on Cloud Computing, vol. 8, no. 1, pp. 246-255, 1 Jan.-March 2020, doi: 10.1109/TCC.2017.2737624.

20.M. A. AlZain, E. Pardede, B. Soh and J. A. Thom, "Cloud Computing Security:From Single to Multi-clouds," 2012 45th Hawaii International Conference on System Sciences, 2012, pp. 5490-5499, doi: 10.1109/HICSS.2012.15

## 2. Overview and Planning

### 2.1 Proposed System Overview

InvoiceInc is aimed to be an open source platform which chartered accountancy firms can leverage/build upon to meet their specific requirements. The platform provides an admin login where the admin has access to all the data and will be able to manage them. There are three kinds of login: Customer, CA employee of the form and Supervisor CA. The customer can only view his/her bills and add his own bills. The employee CA can view and manage the bills of the customers under him. He can also create new expense IDs to categorize the bills into. The supervisor CA can assign customers to the employee CAs manage them and can also view all the bills of all customers. The supervisor also has the ability to add customer accounts. All the bills can be filtered on various criteria and can also be downloaded in various formats (CSV, PDF etc).

Our project will have three major components namely frontend, backend and cloud services used for various specific purposes.

Since we are using the python framework of Django, the frontend will be in the templating language called JINJA and the backend is built in Python using Django modules.

We are using Docker to dockerize our entire project so as to make the software runnable on any system without much configuration. Dockerizing the project will help during system migration or VM migration as the software can run on any VM without negligible initial setup just at the run of one command.

We are using the Google Cloud Platform for our cloud based service.

First of all we are using the Compute Engine - VM Instances to create a vm instance where we can run our dockerized project and serve it on the internet.

Next we are using the Instance Template feature in google cloud to save our main instance as template and generate multiple instances out of it. Once done we can form Instance Groups on google cloud and group all our VM instances into a single group which can then be used to do load balancing of the network traffic using the Network Services - Load Balancing service of GCP.

For storage purposes we are going to use Google Cloud SQL to store our database tables required by our backend.

Other than that we are using VM SSH Manager to manage the ssh keys that can have access to our server and IAM Admin and Roles service to manage who can have access to our google cloud console and their roles regarding what activities they can perform on the console.

## 2.2 Challenges

**2.2.1** Maintaining privacy and security of the bills of the users uploading to the cloud database by the user was difficult.

**2.2.2** Key management for the encryption mechanism of the files is a very complicated process.

**2.2.3** End to end encryption consumes a lot of processing time and causes the system to slow down.

## 2.3 Assumptions

**2.3.1** The admin is the top most authority in the system, and hence has all the access rights.

**2.3.2** There is requirement of only three kinds of users in the system - customers, CA, and CA Supervisor

## 2.4 Hardware Requirements

**A machine with at least 4 GB of RAM and 500 MB of free memory available for the application .Files and other data don't need additional storage because Google Cloud Sql is utilized for the same thing.**

## 2.5 Software Requirements

**Frontend - Jinja**
**Backend - Python and Django**
**Dockerizing - Docker**
**Storage - Google Cloud SQL**

### 2.4 Architecture Specifications

**The application uses Google Cloud Platform for all the cloud based services. Compute Engine - VM Instances are being used to create a vm**

**instance where we can run our dockerized project and serve it on the internet.**

**Each VM instance created have the following specifications:**
**Machine Type - e2-medium**
**Number of CPUs - 2v CPUs**
**RAM - 4GB CPU**
**Platform - Intel Skylake**
**Zone of VM Instance - asia-south1-c**
**Boot Image used on the instance - debian-10-ubuntu**
**Size of VM - 10GB**

The entire storage of the application is implemented using Google Cloud SQL, which is connected to our backend made in Django which allows data flow between the end user and the database.

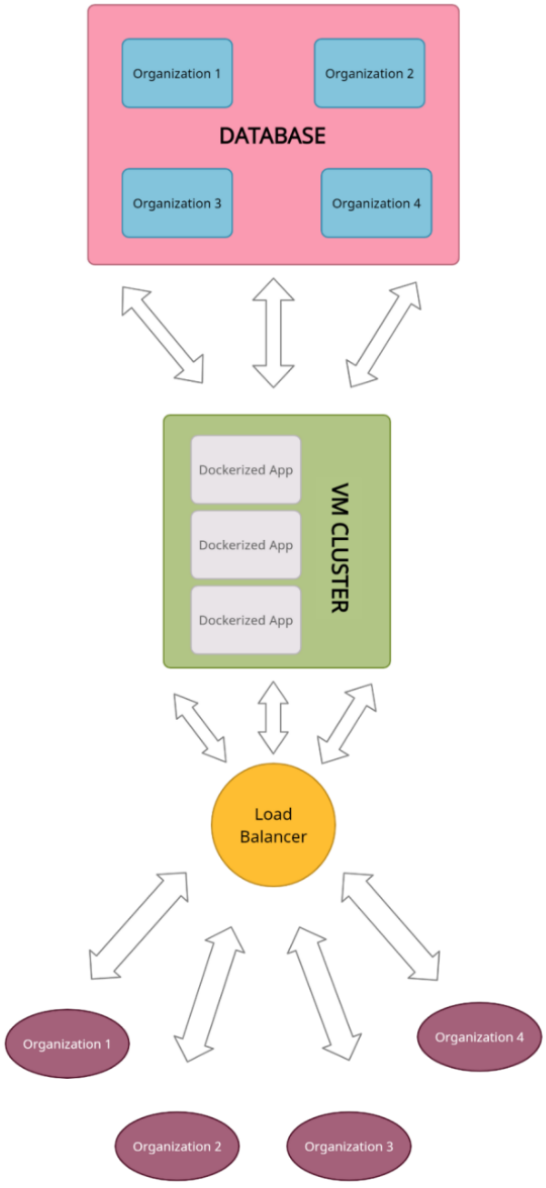
Instance Template feature in google cloud is used to save our main instance as template and generate multiple instances out of it. Once done we can form Instance Groups on google cloud and group all our VM instances into a single group which can then be used to do load balancing of the network traffic using the Network Services - Load Balancing service of GCP.

VM SSH Manager is used to manage the ssh keys that can have access to our server and IAM Admin and Roles service to manage who can have access to our google cloud console and their roles regarding what activities they can perform on the console.

The backend of our project uses the Classic MVC architecture wherein it is divided into three major sections of Models, Views and Controllers. The models hold all the database schemas and their constraints and relations.The views hold all the frontend pages which deal with the logic of how to render data since our project is a server side rendered project. And the controllers consists of functions dealing with all the logic comprising of how to handle data between the user interactions in frontend and the database.

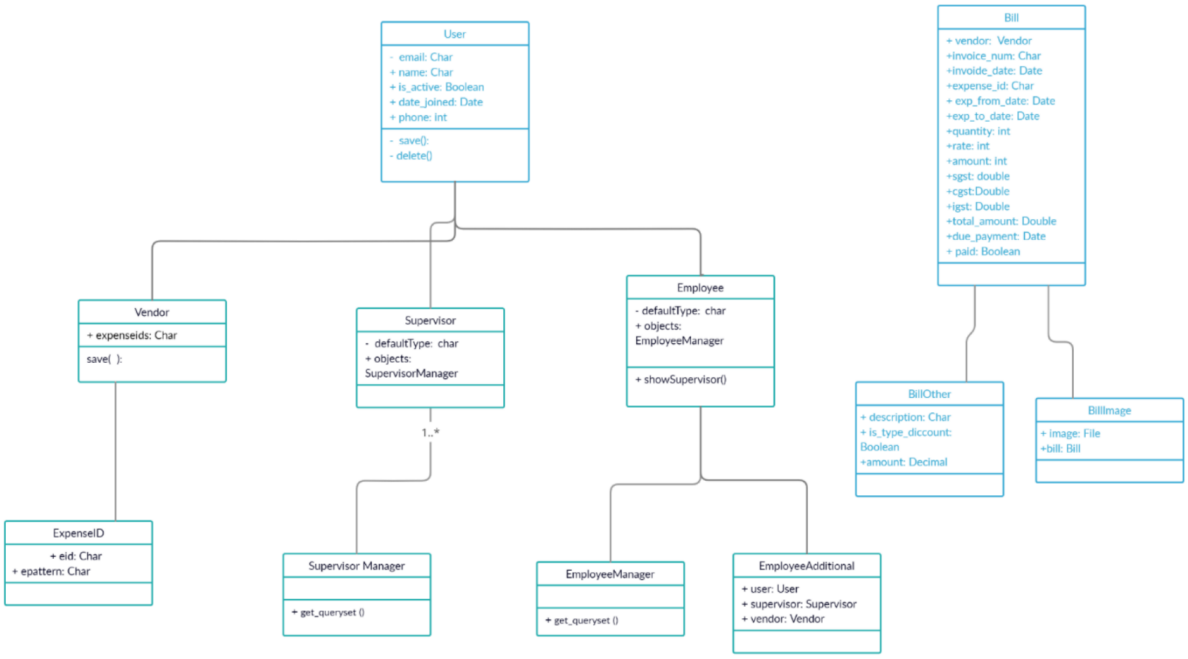
# 3. System Design

## 3.1 High-Level Design

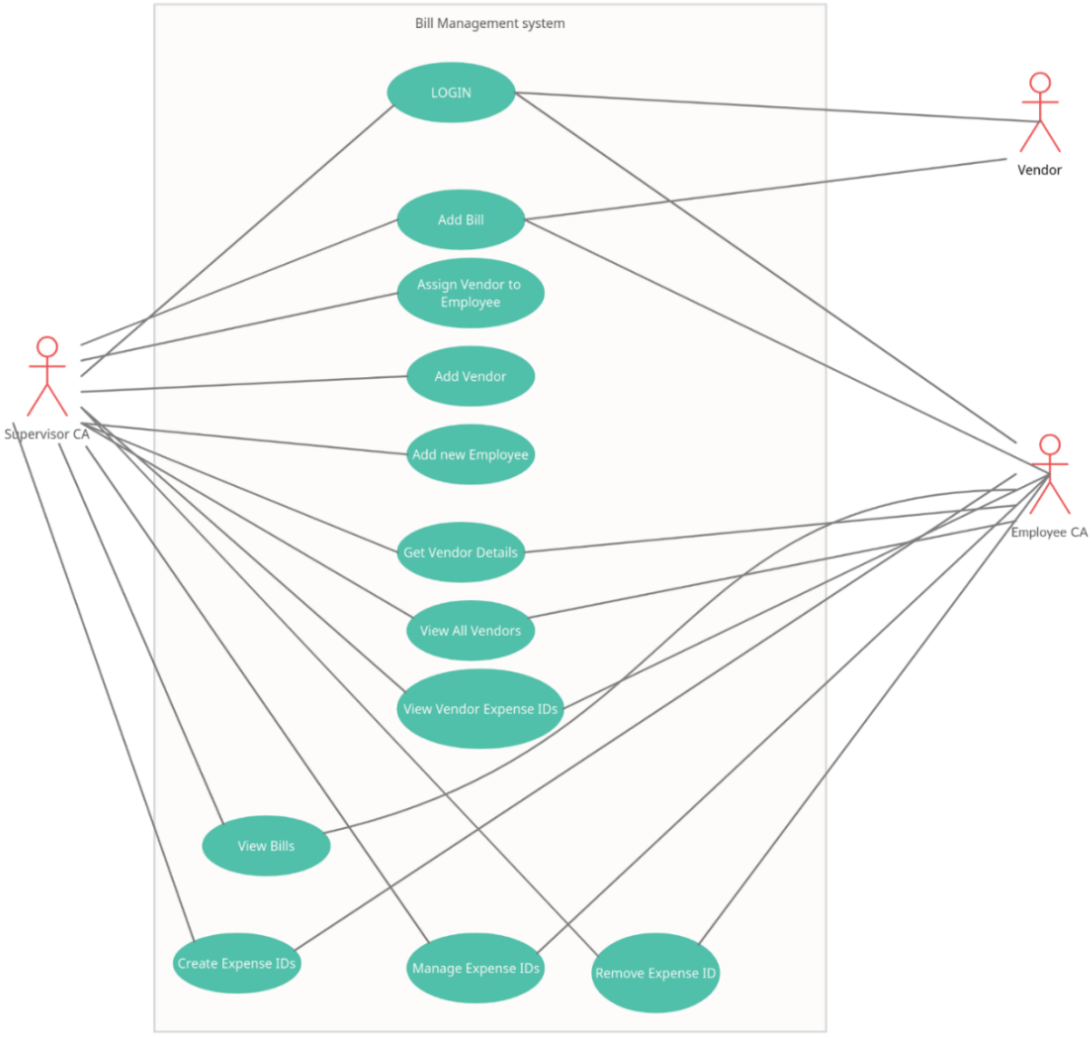


### 3.2.1 Low-Level Design

**The low level design of our application can be well understood using the CLASS and USECASE UML diagrams shown below:**

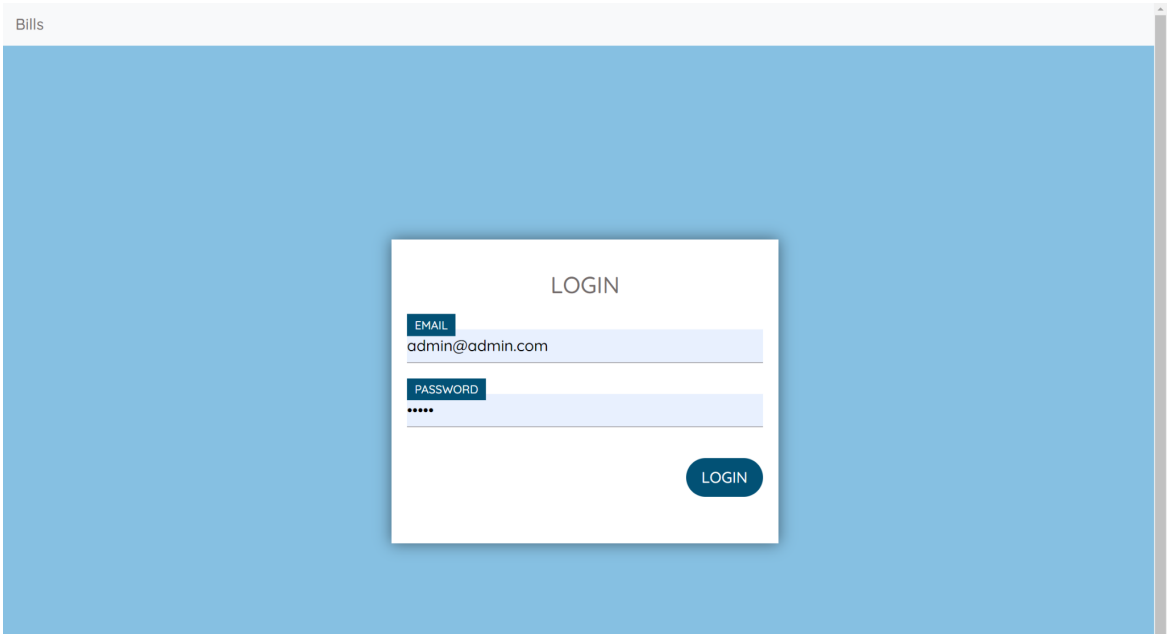


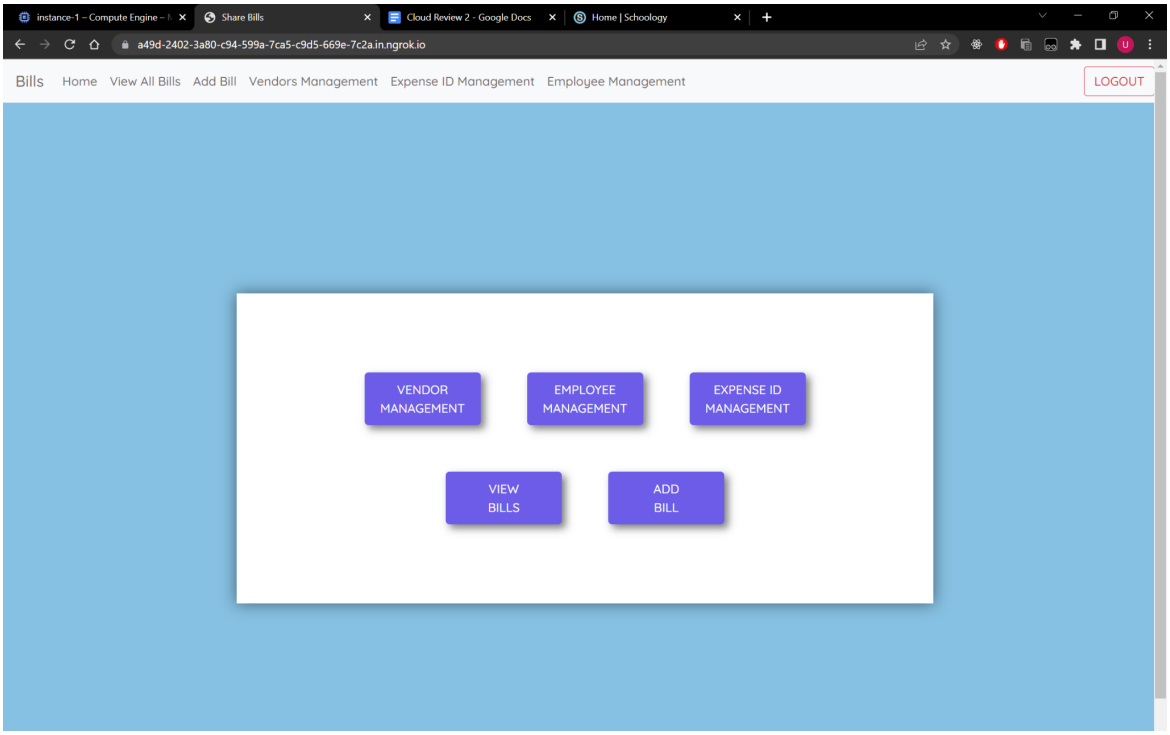
## 3.2.2 Use Case Diagram

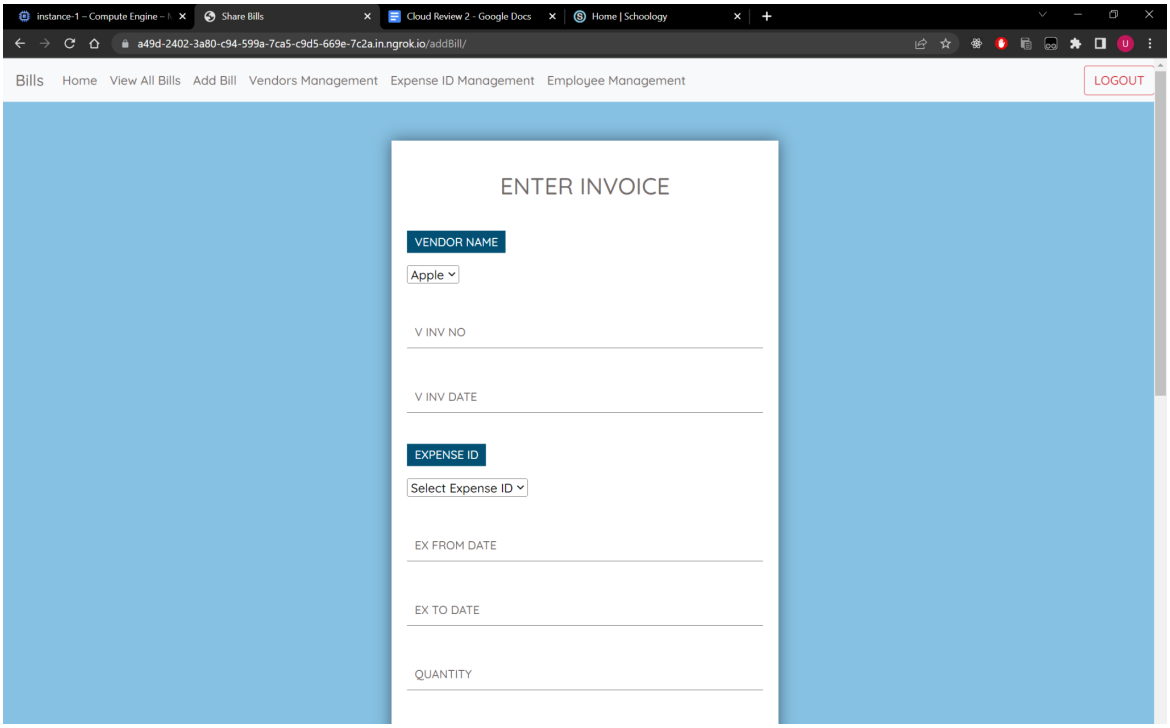


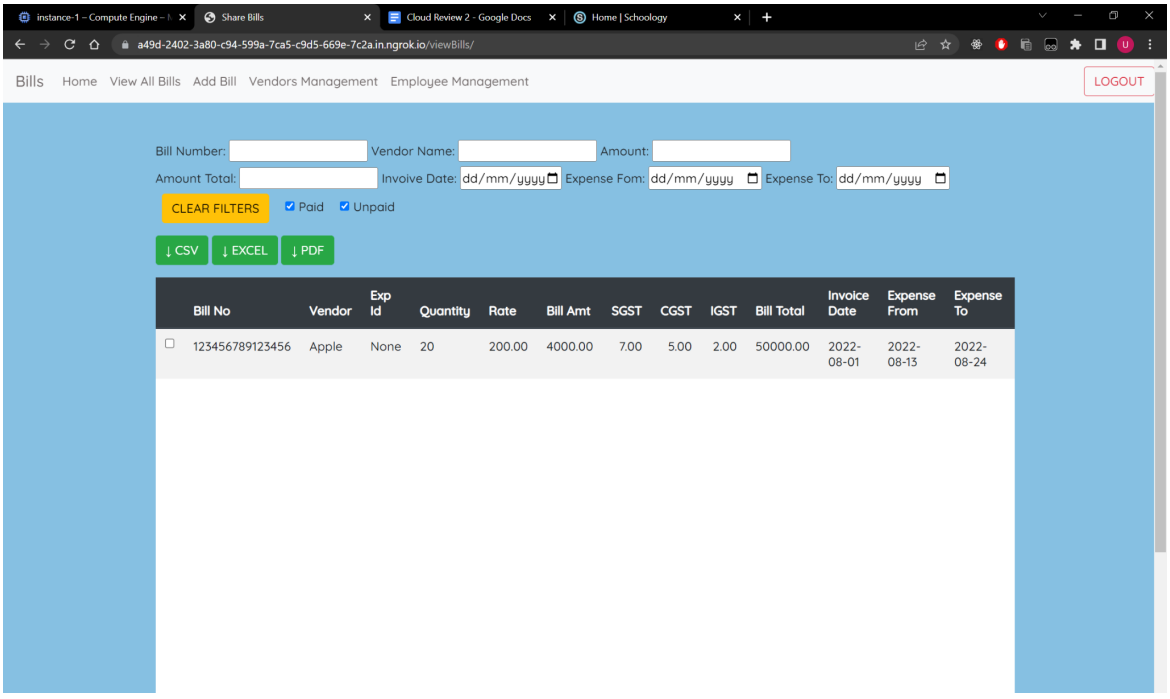
## 4. System Implementation

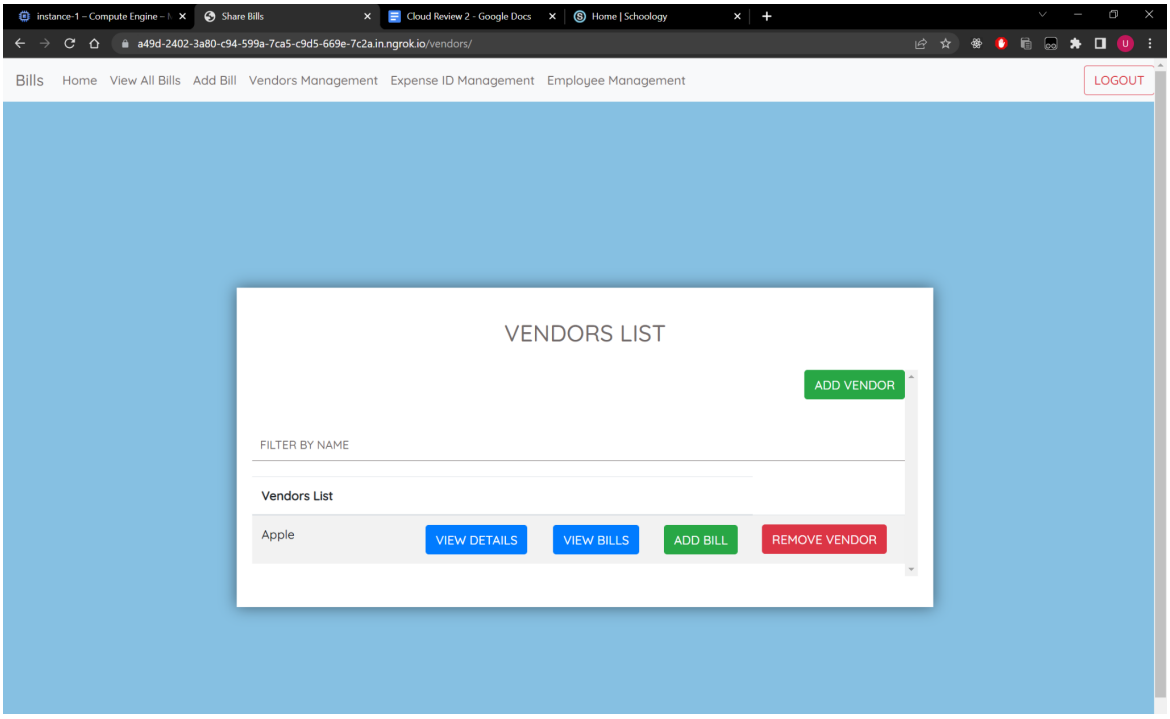
### Screenshots:

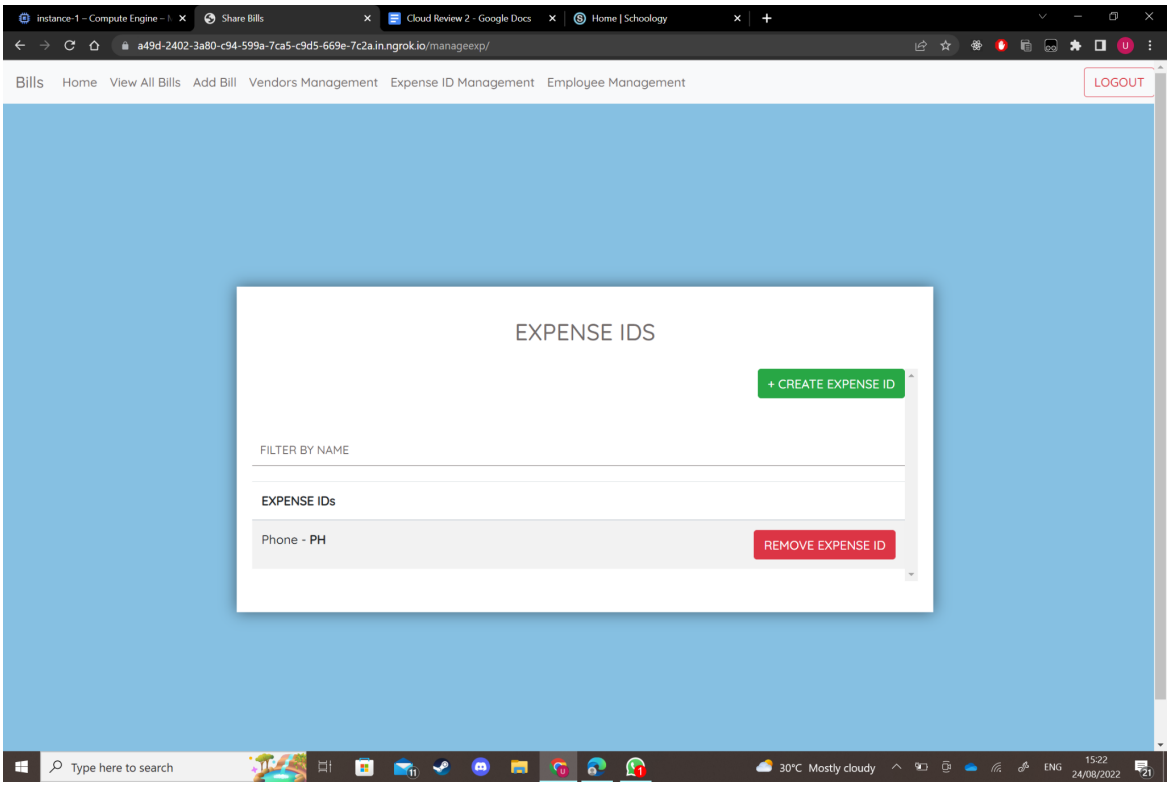


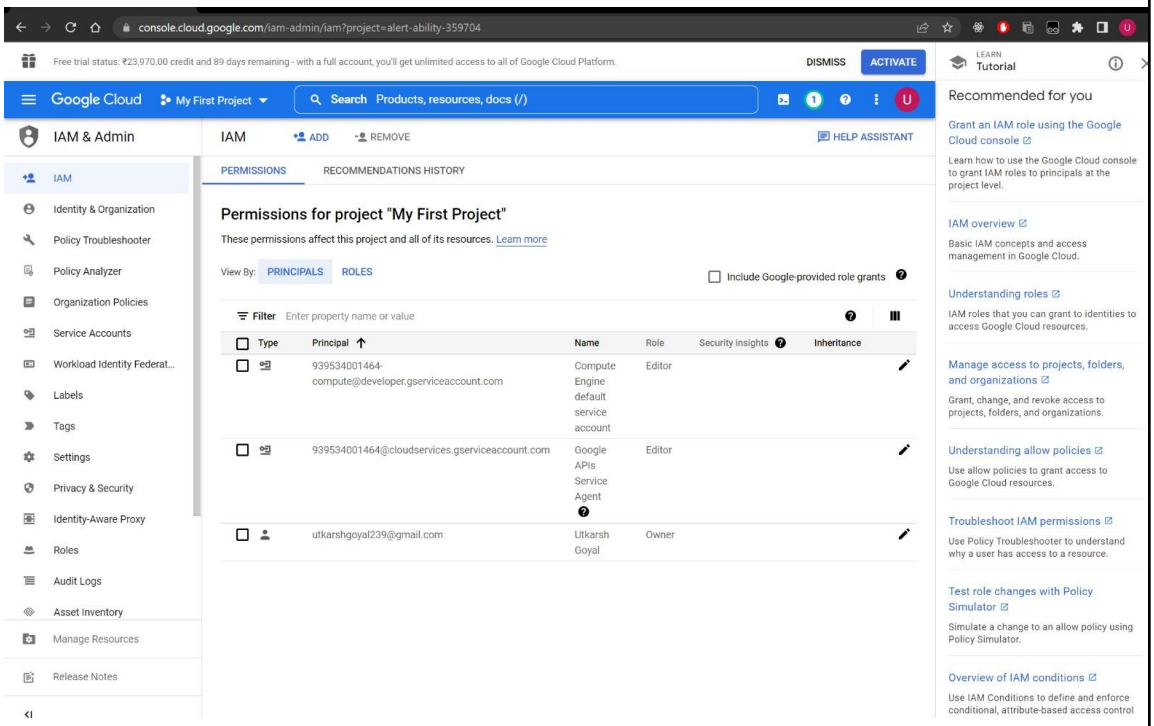

instance-1 – Compute Engine
Share Bills
Cloud Review 2 - Google Docs
Home | Schoology
a49d-2402-3a80-c94-599a-7ca5-c9d5-669e-7c2a.in.ngrok.io/addBill/
Bills
Home
View All Bills
Add Bill
Vendors Management
Expense ID Management
Employee Management
LOGOUT
ENTER INVOICE
VENDOR NAME
Apple
V INV NO
V INV DATE
EXPENSE ID
Select Expense ID
EX FROM DATE
EX TO DATE
QUANTITY


instance-1 – Compute Engine
Share Bills
Cloud Review 2 - Google Docs
Home | Schoology
a49d-2402-3a80-c94-599a-7ca5-c9d5-669e-7c2a.in.ngrok.io/viewBills/
Bills
Home
View All Bills
Add Bill
Vendors Management
Employee Management
LOGOUT
Bill Number:
Vendor Name:
Amount:
Amount Total:
Invoive Date:
dd/mm/yyyy
Expense Fom:
Expense To:
CLEAR FILTERS
Paid
Unpaid
↓ CSV
↓ EXCEL
↓ PDF
Bill No
Vendor
Exp Id
Quantity
Rate
Bill Amt
SGST
CGST
IGST
Bill Total
Invoice Date
Expense From
Expense To
123456789123456
Apple
None
20
200.00
4000.00
7.00
5.00
2.00
50000.00
2022-08-01
2022-08-13
2022-08-24

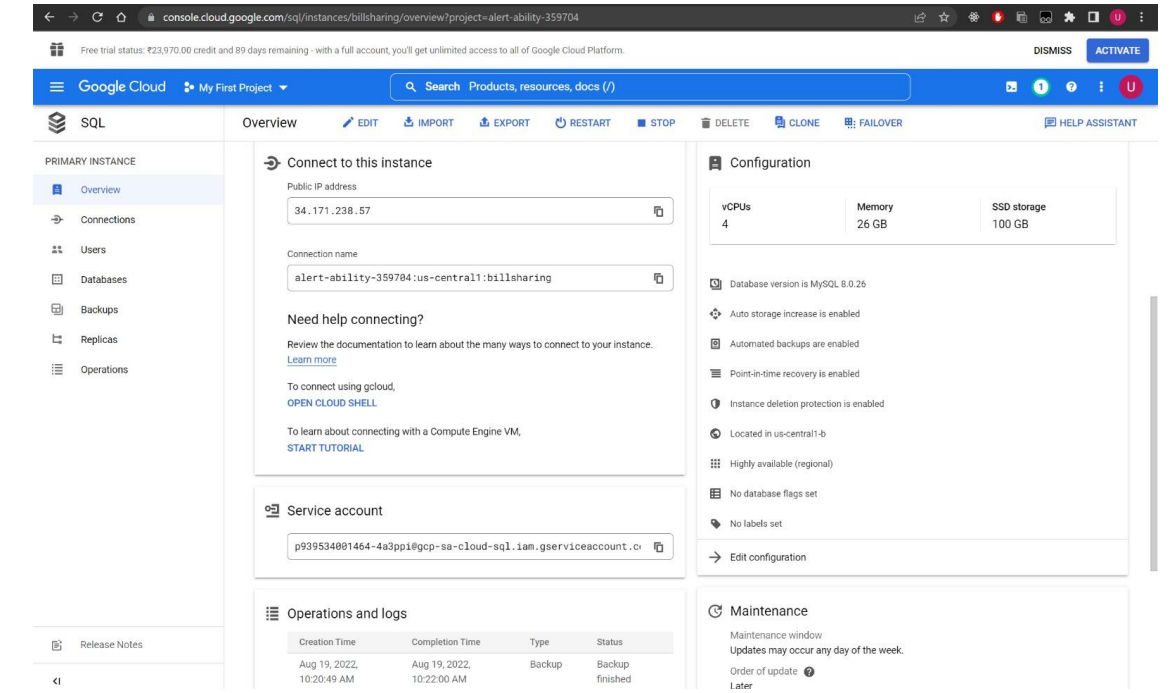

instance-1 – Compute Engine
Share Bills
Cloud Review 2 - Google Docs
Home | Schoology
a49d-2402-3a80-c94-599a-7ca5-c9d5-669e-7c2a.in.ngrok.io/vendors/
Bills
Home
View All Bills
Add Bill
Vendors Management
Expense ID Management
Employee Management
LOGOUT
VENDORS LIST
ADD VENDOR
FILTER BY NAME
Vendors List
Apple
VIEW DETAILS
VIEW BILLS
ADD BILL
REMOVE VENDOR

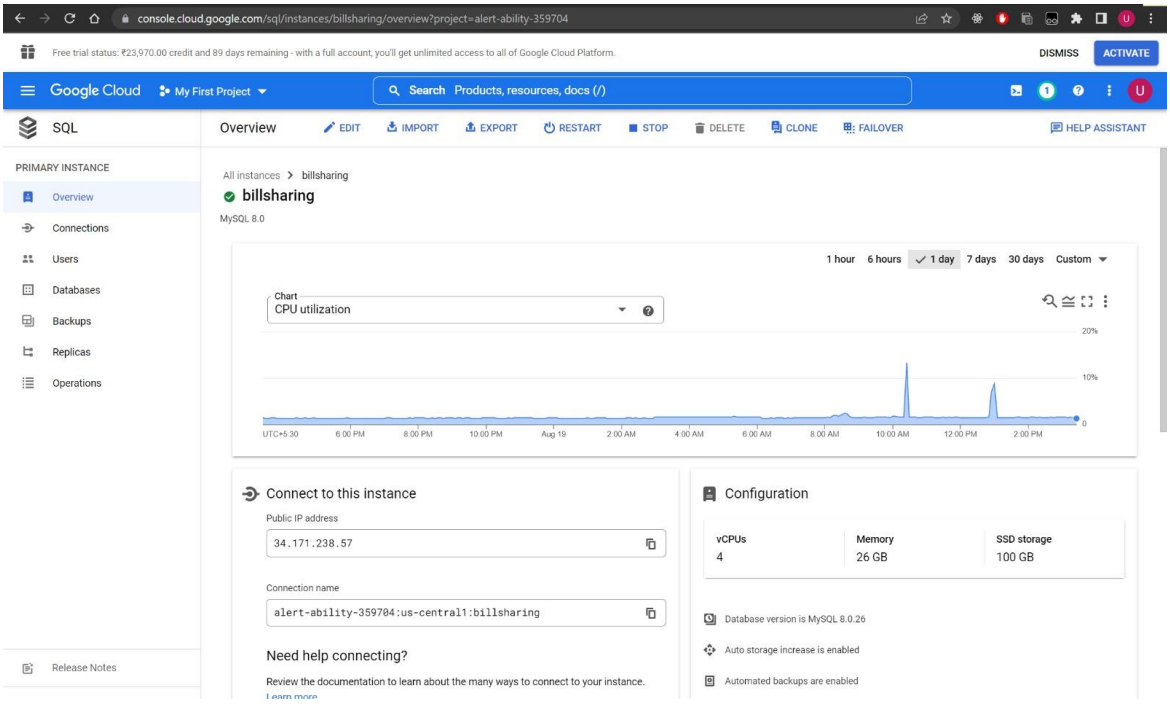

instance-1 – Compute Engine
Share Bills
Cloud Review 2 - Google Docs
Home | Schoology
a49d-2402-3a80-c94-599a-7ca5-c9d5-669e-7c2a.in.ngrok.io/manageexp/
Bills
Home
View All Bills
Add Bill
Vendors Management
Expense ID Management
Employee Management
LOGOUT
EXPENSE IDS
+ CREATE EXPENSE ID
FILTER BY NAME
EXPENSE IDs
Phone - PH
REMOVE EXPENSE ID
Type here to search
30°C Mostly cloudy
ENG
15:22
24/08/2022

# IAM Roles Manager Permission Management on GCP

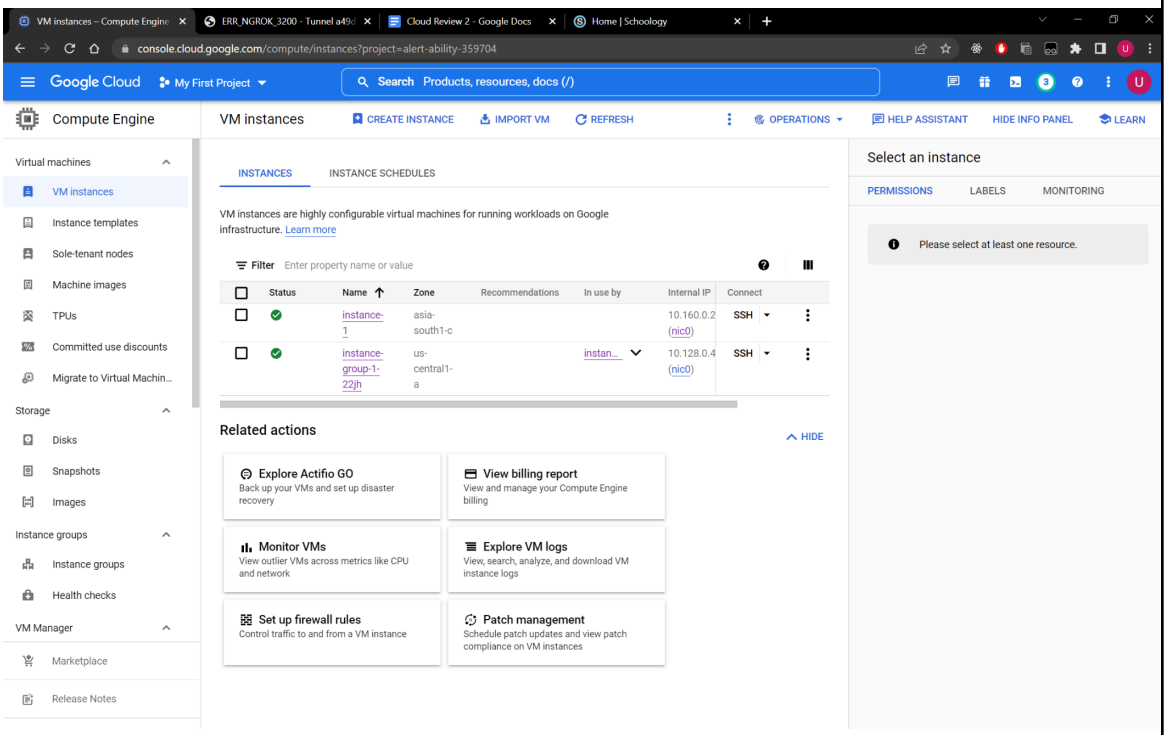


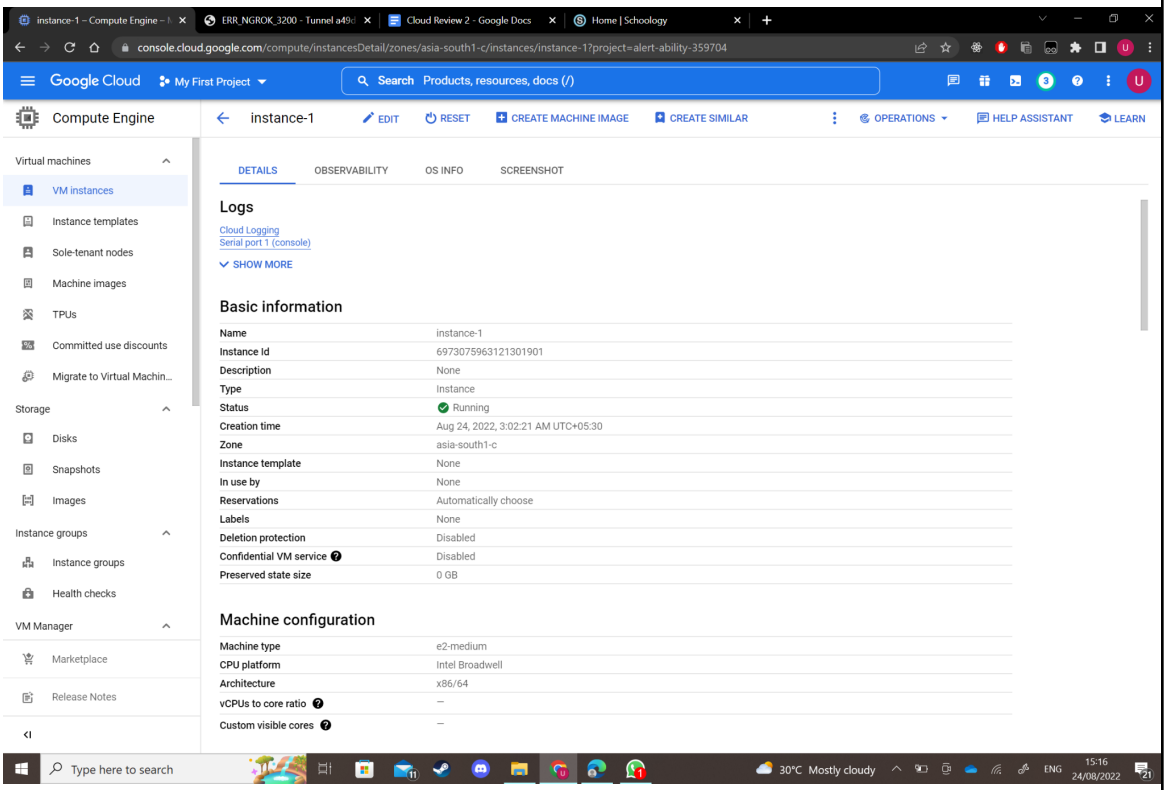
# Mysql template created



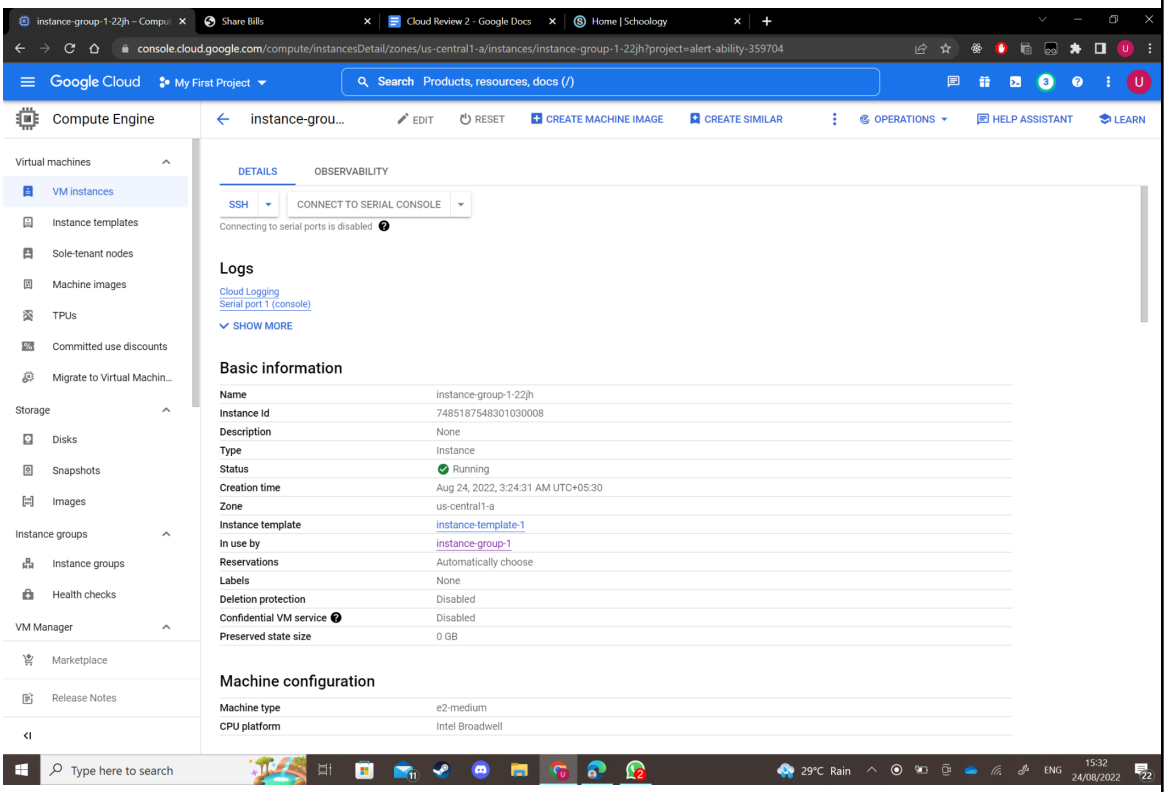



# VM Instances created on GCP

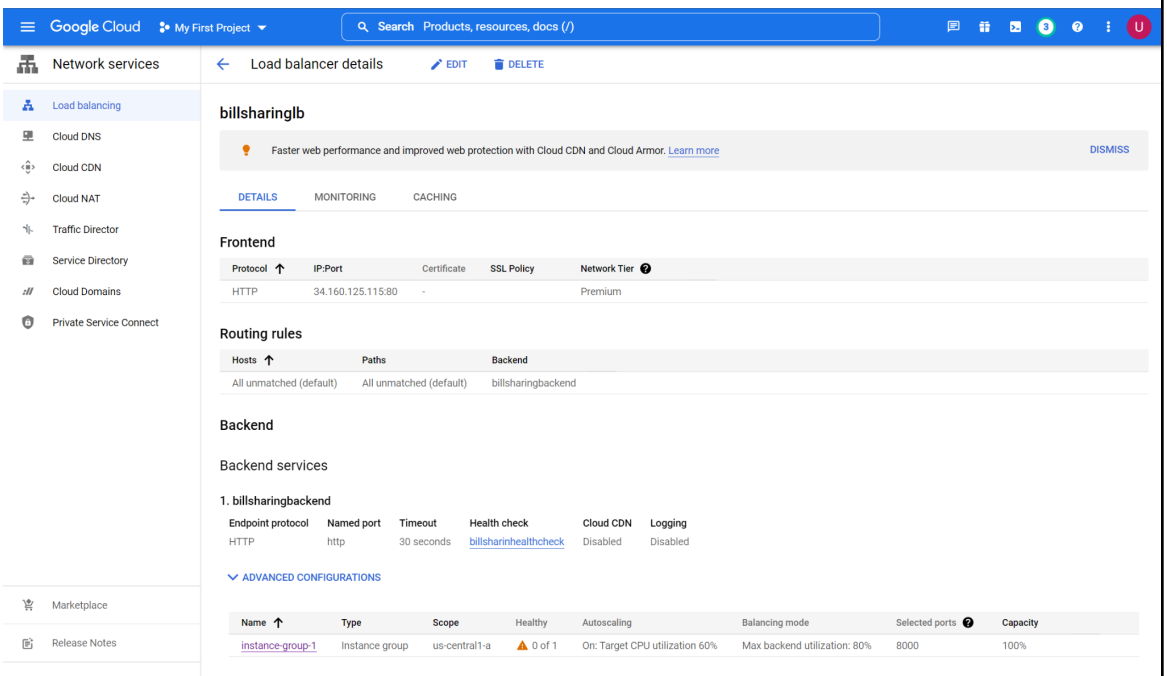


# Instance Group Configuration




# Load Balancer Configuration



**5. Results and Discussion:**

We were able to create an open source accountancy application to help charted accounts in managing the bills of their customers along with the security of google cloud and encryption algorithms, which prevent any possible tampering of the data uploaded to the cloud servers by any party. Moreover our application is deployed over several Virtual Machines, managed by a load balancer, which enables it to handle a high amount of traffic at any point of time.

**6. Conclusions and Future Scope:**

Our application is the only open source solution currently available to help chartered accountants manage bills in a secure way. In the future the scope of the application can also be extended to organizations of accountants who work in a distributed manner without the need of any super admins. Encryption of the images of the bills can also be implemented in the future using some image encryption algorithms like AES, DES etc. Further using some visualization techniques we can also help the customers understand their spending patterns and recommend ways in order to help them in tax savings.

| | |
|---|---|
| | **B.Tech (Information Technology) SUMMER SEMESTER TERM 5 - 2022**<br>**ITE 3007 : Cloud Computing and Virtualization**<br>**19.08.2022** |
| **Title: AccountantPro:CA platform to manage transcripts and delegate employees to clients** | |
| **Team Name: Deadpool** | |
| **Project Team: Deadpool** | |

| S.No | Register Number | Student Name | Signature | Guided By |
|---|---|---|---|---|
| 1 | 19BIT0402 | Utkarsh Goyal | | **Dr. Nadesh R.K** |
| 2 | 19BIT0354 | Dipesh Balani | | |
| 3 | 19BIT0225 | Aayu Ojha | | |

**Team Member(s) Contribution and Performance Assessment**

| Components | Student 1 | Student 2 | Student 3 |
|---|---|---|---|
| **Implementation & Results -(20)** | | | |
| **Contributed fair share to the team project -(05)** | | | |
| **Documentation without Plagiarism -** | | | |

| | | | |
|---|---|---|---|
| **(20)** | | | |
| **Q & A - (05)** | | | |
| **Student Feedback**(Student Experience in this Course Project) | | **Evaluator Comments** | |
| **Name & Signature of the Evaluator**<br>**(Dr.Nadesh R.K)** | | | |