Architecture Used for Data Management in Cloud Based System

[1]Hira Lal
[1]Research Scholar in Dr. A. P. J. Abdul Kalam University,Indore, Madhya Pradesh
[2]Dr. Arpana Bharani
[2]Assistant Professor, Department of Computer Science, Dr. A. P. J. Abdul Kalam University,Indore, Madhya Pradesh

ABSTRACT
A focus on the most efficient method of data management is required in this case. When it comes to cloud data storage, Big Table and Data Store are two options that work in distinct ways. In order to achieve our goals, it must make use of the appropriate technologies.One of the newest technologies to gain pace in recent years is cloud computing.
Rather of relying on local computers, data and software are stored and sent via the internet with this architecture. Cloud resource management is a difficulty since demand is always changing and there are so many concurrent users. Cloud resource management is complicated by server sprawl, a problem that is all too typical in data centres. Overcrowding on servers results in underutilization because of the need for more resources that aren't justified by workload.In this thesis, effective resource management may be used to successfully consolidate servers in cloud data centres. All nodes are housed inside their own virtual machines as a result of the thesis' design of a cost matrix technique. It is possible to maximise the memory of each node by utilising this method, which leaves the rest of the nodes free for other processes to utilise. Virtualization technology is used to allocate and relocate virtual computers from real ones in order to reduce downtime. It was able to reduce server costs and downtime by using the best fit concept. To avoid wasting resources, virtual machines are placed in memory that has spare space and then relocated when the memory is full. The load-balancing hotspot algorithm is also becoming better.
KEYWORDS: Data Management, Cloud Based System, cloud data storage, Cloud resource management

INTRODUCTION

The client tier layer or IoT devices were used to collect the first data shown in Figure 1. To describe IoT devices or the client tier, "layer" might be used instead of "tier." When a result, the quality of service will be improved as data is routed via Low-latency location awareness is provided by fog devices. SLG or SG will receive data for processing or management. There are a range of operational and energy measures that make up a smart grid, the use of energy-efficient technologies such as smart metering, energy-efficient appliances, and renewable energy. There is a heavy reliance on electronic power conditioning and energy production and distribution control in the smart grid. The terms "smart local grid" and "smart grid layer" are used inter changeably here. These are the information will be transferred to the cloud after this step. Local storage is provided by Smart Local Grid, while a central semi-permanent computing stage is provided by Smart Grid. Initially in our original statement, the data will be generated by either a client or an IoT device. The data generated as a consequence of these devices is referred to as the Internet of Things (IoT). Data from the Internet of things is sent to a fog computing layer for processing or management. To create a smart

grid, fog computing was included. Once the data has been processed, we gather the new data and remove the old data. Local storage area is used to keep the most recent copies of the data. These locally stored data are then sent to a central storage location, where they are available to all users. Finally, these new data are permanently saved in the cloud. Consequently, of the fast growth of smart devices and smart infrastructures, IoT devices generate the majority of the big data. It is critical, thus, that these IoT sensors be organised properly. Cloud computing is a permanent processing area, while fog computing is a semi-permanent stage. In order to provide services in real time. The smart grid, whether it's a national or a municipal one, is critical.

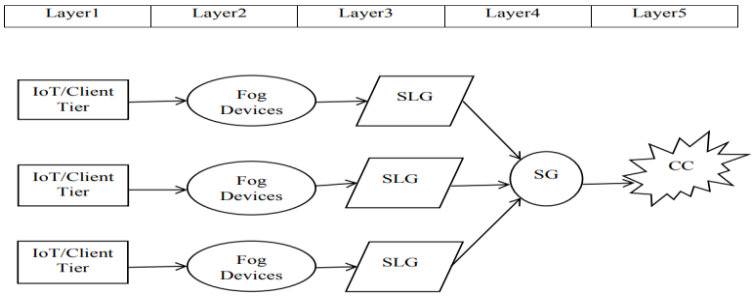

Figure. 1: Big Data Management Architecture

DATA MANAGEMENT

I've employed the following parameters in my design to handle or process data for the smart grid or the smart local grid:

The safety and privacy of your personal information is extremely important to us. It's possible to number nodes in a hierarchical structure by utilizing BFS (Breadth First Search). As a consequence, it can rapidly pinpoint the cause of the problem without having to walk to every node.

Cost-cutting: Only maintenance expenditures are necessary if the architecture can be finally rectified once and for all.

Privacy and data storage: Real-time services may benefit from the employment of distributed systems for data storage and privacy.

Fog infrastructure may be dynamically priced, enabling customers to retain a fixed rate by expanding or lowering the number of nodes.

REDUCING THE VOLUME OF BIG DATA BY PROCESSING AND CONTROLLING THE DATA

Layers of fog computing are shown in threes in Figure2. IoT data has been collected by Sublayer1. Sublayer 2 is then filtered, and any unnecessary data is removed. This new information has been gathered by Sublayer3 and is now available for use. There should be a record of all the metadata, key data in terms of conditioning, relevant data, and timestamps associated with the new data. Fog devices are sublayer 1 in our concept, whereas Smart Grid and Smart Local Grid are sublayers 2 and 3 respectively.

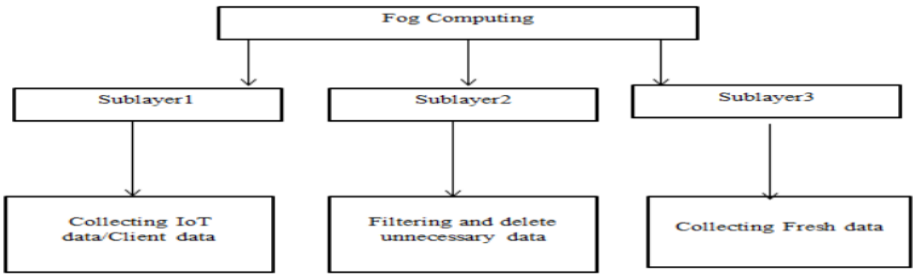

Figure. 2: Fog Computing Sub layer to Find Fresh Data from Big Data

REAL TIME SERVICE DELIVERY

For latency-sensitive applications, it may simply supply real-time services utilizing fog computing. Fog devices are making it easier to keep track of where you are. One of the greatest challenges with fog devices is that they have incredibly little memory.

Third-party memory management, known as a smart local grid and a smart grid central storage has been introduced to our architecture. When real-time services are needed, local storage is preferable as it can immediately offer the data necessary. Using Semi-permanent energy storage in the form of a Smart Grid or Smart Local Grid has been a prevalent strategy in fog computing. Figure 3 displays the aforementioned. To give real-time services, previous data should be kept in the cloud.

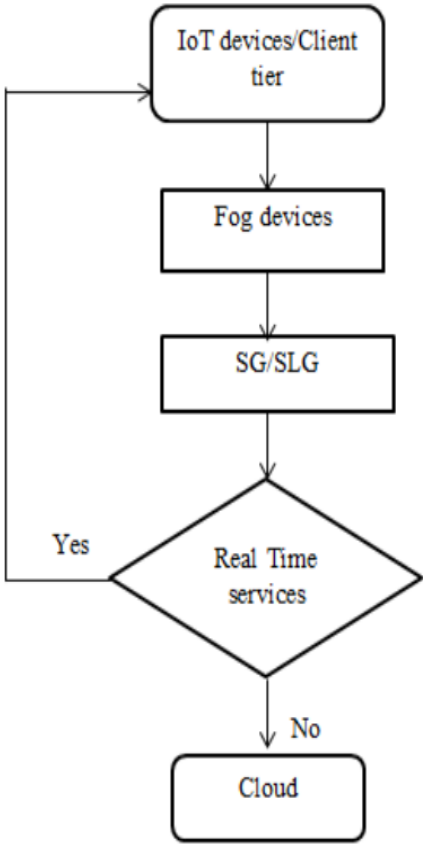

Figure. 3: Flow Chart for Real Time Service from Fog Computing Layer

FRAMEWORK ARCHITECTURE

The architecture of the proposed integrated framework is as depicted in Figure 4.

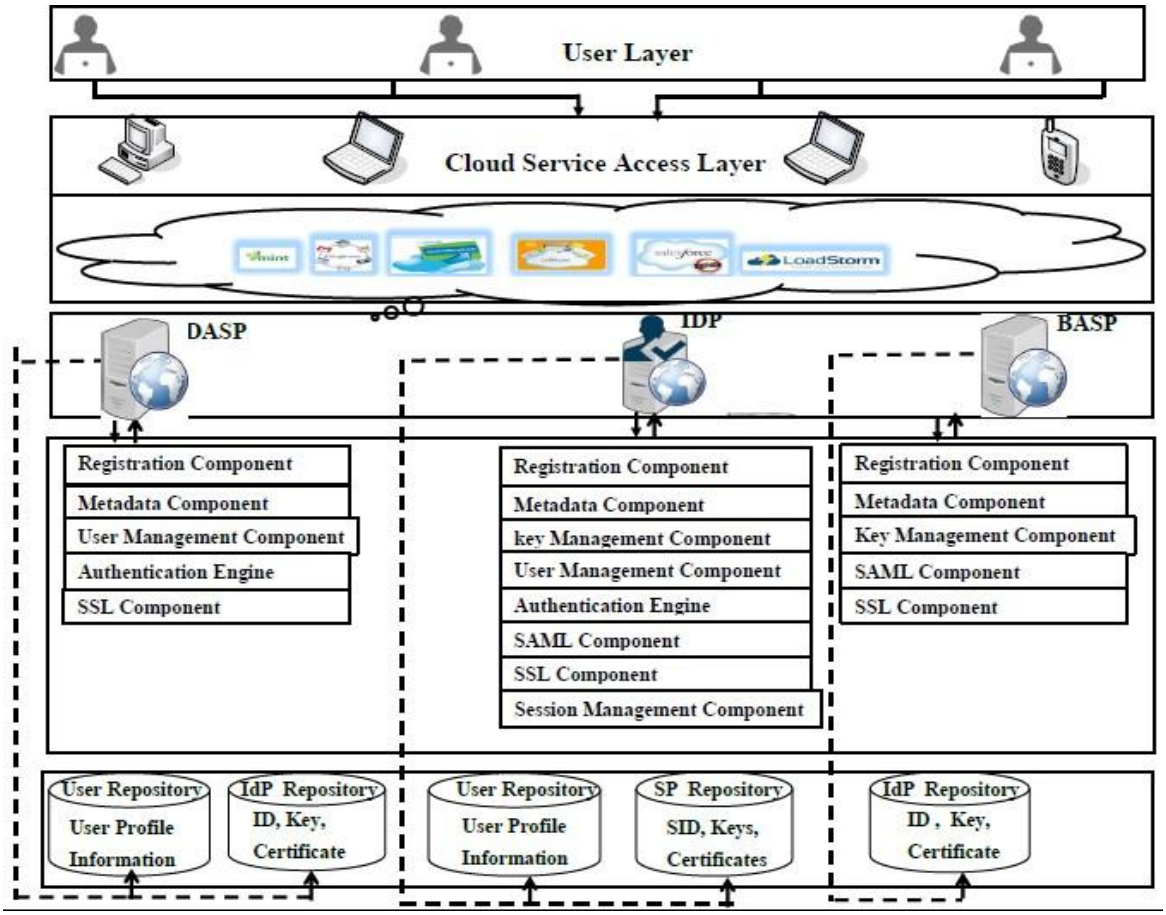

Figure 4 Framework Architecture

The role of the participants of the framework and the functioning of the components are discussed in the following paragraphs.

Users: Users access the services offered by numerous Service Providers following registration and authentication.

It is important for a user to register with the Identity Provider (IdPregistration)'s server in order to utilize the services of registered service providers. A crypto-token or mobile token comprising the user's credentials and the IdP's secret key will be issued to the user after registering at the IdP. There will also be a list of service providers that the user may access. Both Identity Providers and Service Providers do not maintain a user's password. The Crypto-token or the mobile-token comprises a password variant. The user's crypto-token/mobile-token is authenticated by checking the security parameters provided in the token during the authentication process. The user's password and a crypto-token/mobile token are the two authentication pieces in this integrated design. In order to access direct authentication-based services, the user must authenticate to each service provider. It is not essential for the user to authenticate more than once per session if the service provider employs brokered authentication.

Service Providers: Registration servers for Identity Providers should be used to register Service Providers that are a member of the framework (IdP). Direct Authentication Service Providers (DASP's) and Brokered Authentication Service Providers (BASP's) are two categories of service providers.

***Registration Component of DASP*:** It is the registration component of BASP that controls service providers who employ brokered authentication to register themselves. They must offer a server ID, URL, a short description of what they do and their chosen authentication option as "Brokered authentication" in order to be approved as a BASP. The master key of IdP is securely provided to BASP at the completion of the registration process. BASP and the IdP both create a shared key with the aid of this private key. BASP utilizes this shared key to exchange the session key established by IdP and the user at the time of each authentication session utilizing this shared key.

***Registration Component of BASP*:** This component manages IdP information, such as the IdP's unique ID and URL. In order to contact with IdP, this information must be submitted to them.
BASP's Metadata Component: This component maintains track of IdP information, such as its unique ID and URL. If you wish to perform a SAML authentication request to an IdP and verify the IdP's authentication response, you'll need this information from BASP.

***User Management Component of DASP*:** Security Assertion Markup Language (SAML) is used to transfer the user's authentication information between the Identity Provider and BASP during the authentication process (SAML). All of BASP's security features, including SAML, are handled by this component. In the request, the SP (unique ID) who made the request is checked against the metadata information maintained about the SP's by the IdP, the unique ID of the IdP, the Assertion Consumer Service (ACS) URL, which is the site to which IdP's SAML authentication response is delivered.

***SAML Component of BASP*:** Authentication and authorization statements, a unique assertion identifier, an issue instant (the time when the assertion was created), and the issuer name (the IdP name) are all included in the SAML assertions (responses) sent by the IdP, and this data is checked against metadata about the IdP maintained by the SP. The SP checks the user's SAML answers and, if successful, allows access to resources and services.

***Authentication Engine of DASP*:** This is how DASP's two-factor authentication protocol is carried out by the authentication engine. The authentication server of these service providers will authenticate users who seek to utilize their services. Server-side authentication is performed by the module of the server's authentication server. In order to proceed, the user management component will be queried by the authentication component to determine whether the user has been registered. The specified authentication protocol and crypto-token/mobile-token are employed as the two authentication components by the authentication module. After completing the mutual authentication procedure, a session key is created and used to safeguard ongoing communications.

***Key Management Component of BASP*:** Management of BASP is a vital component of this system. By employing BASP's key management component, the service provider's master key may be securely controlled. During the brokered authentication process, this component also securely controls the session keys created between the user and the IdP, and communicated to BASP by the IdP at the end of the session.

***Secure Sockets Layer* (*SSL*) *Component of DASP*/*BASP*:**, the user's credentials should be delivered to the Identity Provider in a secure way. Using an SSL connection assures the secrecy of information exchanged over the network. The SSL component is responsible for establishing and maintaining an SSL connection.

***User Repository of DASP*:** Repository maintains track of all the persons that have signed up for the system.
Information about the Identity Provider (IdP), such as its unique ID, Domain name, Digital Certificate of IdP and the terms and conditions of its Business Agreement (BA), is held in this repository.

Identity Provider: IdP is a trustworthy entity that offers registration and authenticating services on demand. Single Sign-On may be made simpler by setting the IdP's digital certificate and the SAML protocol.

Identity Provider: Participants in the proposed framework will be service providers that have registered with the identity provider to supply cloud applications and services. They should utilize IdP's registration server and submit a unique ID for their servers, as well as the service provider's URL (URL for direct authentication), a short explanation (description) of what they offer, and their chosen authentication method (direct or broker) (direct or broker). DASP will acquire an authentication module from the IdP following registration. DASP obtains a secure copy of the IdP's master key during registration, and this key is then used to verify an authentication parameter during the 2-factor authentication protocol execution with the user. It is this key that is exploited by BASPs to construct an IdP/BASP shared key.

***User Management Component*:** The user management component of IdP Users controls the profile information of users registered with IdP.

***Metadata Component*:** The metadata component is in charge of keeping track of all DASPs and BASPs that have been registered with IdP. DASP's that should be informed of registered users' profile information are indicated by metadata component upon registration. The metadata component lets the IdP verify the origin of the authentication request during Brokered Authentication.
***Authentication Service*:** Authentication as a service is offered by IdP, which validates the identities of users who have been routed to it through BASPs. The IdP gets a SAML response from the authentication server of the IdP after the two-factor authentication protocol has been exchanged with the user.

***SAML component*:** During the Brokered Authentication process, the IdP's SAML component should be configured to authenticate SAML requests received from BASPs and to create SAML replies in compliance with the standards of SAML v2.0.

***Key Management Component*:** The IdP's Key Management Component produces and maintains shared keys with the BASPs. This component is also responsible for securely delivering the session keys established by the IdP and user during the authentication process to BASP.

***SSL Component*:** During the login and authentication phase, all communication between the user and the server should be secured using SSL. Using an SSL connection assures the secrecy of information exchanged over the network. Setting up and maintaining a secure SSL connection is the major duty of the SSL component.

***Session Management Component*:** To ensure that a user authenticates once, throughout a session, for many services, the Session Management Component of the IdP is necessary (SSO).

***Service Provider Repository*:** This repository provides information about SPs, such as their unique ID, domain name, digital certificates, preferred mode of authentication, sort of services given by SP, Business Agreement (BA) terms and conditions, etc.

INTEGRATED AUTHENTICATION MODEL FOR CLOUD

In the integrated model for authentication of users, direct authentication service providers and Brokered authentication service providers should be registered with the registration server of IdP. Authentication of user will be done by the Authentication Server (AS) of the DASP in the case of direct authentication, and Authentication Server (AS) of the IdP would authenticate the user in the event of brokered authentication. The registration and authentication procedure flow are as depicted in Figure 5.

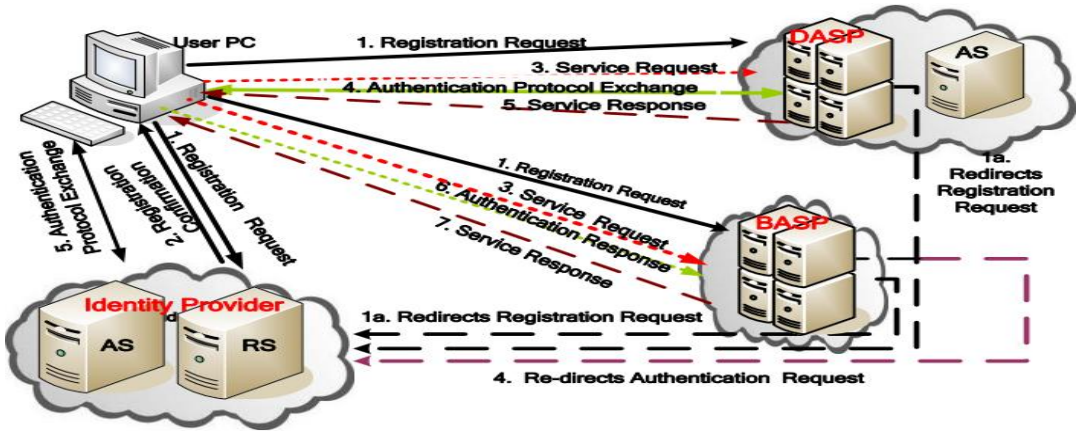

Figure 5 Registration and Authentication Process Flow for Framework

CONCLUSION

Existing information may be analysed and synthesised to produce new knowledge, whether implicit or explicit. Having the freedom to pursue new paths of study is critical. In spite of the rising popularity of data management, individuals still need a way to express their ideas and insights. In addition, this research will showcase the most recent and creative cloud data management innovations. Petabytes of data management, massively parallel query execution, analytic analysis tools, and real-time query processing are some of the issues that cloud data management brings. In addition, this article examines how public, private, and hybrid cloud systems are really utilised in practise. Access, ownership, and location are all factors to consider when choosing a cloud service because they provide enterprises complete control over their data, services, resources and infrastructure.

Private clouds are the best option for corporations". The paucity of research in this field is a problem.A hybrid cloud management approach has been created because public clouds are dangerous and unsuited for commercial use.

Many organisations no longer have control over data, services, and technology. By combining private and public cloud resources, a hybrid cloud may be established. In spite of the limitations of accessibility, analytical processing, and query processing, new research is conceivable in this domain. Hybrid cloud management should also be considered. Data quality may be harmed by merging, aggregating, trading, or mixing data in many ways. There has to be a lot more study done in this area since there isn't any.

According to several studies on cloud-based data management, operational, transactional databases will almost certainly be replaced by cloud computing platforms for decision support systems, tasks, and application-specific data marts in future. Cloud concepts including reliability, availability, security, and privacy are central to this research. An overview of the current state of cloud-based data management and the areas of research that are most important to practitioners is presented in this study understanding the history and current state of cloud-based data management can aid in this endeavour. Unresolved challenges in cloud data management, such as data security and privacy for private and personal clouds, remain.

REFERENCES

1. Brous, P., Janssen, M., & Herder, P. (2019). Internet of Things Adoption for Reconfiguring Decision-Making Processes in Asset Management. Business Process Management Journal, 25(3), 495-511.
2. Boos, D., Guenter, H., Grote, G., & Kinder, K. (2013), Controllable Accountabilities: The Internet of Things and its Challenges for Organisations. Behaviour & Information Technology, 32(5), 449-467.
3. Celesti, A., Fazio, M., Villari, M., &Puliafito, A. (2012). Virtual Machine Provisioning through Satellite Communications in Federated Cloud Environments. Future Generation Computer Systems, 28(1), 85-93.
4. Chiregi, M., &Navimipour, N. (2016). A New Method for Trust and Reputation Evaluation in the Cloud Environments Using the Recommendations of Opinion Leaders' Entities and Removing the Effect of Troll Entities. Computers in Human Behavior, 60, 280-292.
5. Daraghmi, E.Y., & Yuan, S. (2015). A Small World based Overlay Network for Improving Dynamic Load-Balancing. Journal of Systems and Software, 107, 187-203.
6. De Falco, I., Laskowski, E., Olejnik, R., Scafuri, U., Tarantino, E., &Tudruj, M. (2015). Extremal Optimization Applied to Load Balancing in Execution of Distributed Programs. Applied Soft Computing, 30, 501-513.
7. Dodonov, E., & de Mello, R. (2010). A Novel Approach for Distributed Application Scheduling based on a Prediction of Communication Events. Future Generation Computer Systems, 26(5), 740-752.
8. Gangwar, H., Date, H., & Ramaswamy, R. (2015). Understanding Determinants of Cloud Computing Adoption Using an Integrated TAM-TOE Model. Journal of Enterprise Information Management, 28(1), 107-130.
9. Goscinski, A., & Brock, M. (2010). Toward Dynamic and Attribute based Publication, Discovery, and Selection for Cloud Computing. Future Generation Computer Systems, 26, 947-970.

10. Hoefer, C.N., &Karagiannis, G. (2010). Taxonomy of Cloud Computing Services. IEEE Globecom Workshop on Enabling the Future Service-Oriented Internet.
11. Kumar, S. (2022). A quest for sustainium (sustainability Premium): review of sustainable bonds. Academy of Accounting and Financial Studies Journal, Vol. 26, no.2, pp. 1-18
12. Allugunti, V.R. (2019). Diabetes Kaggle Dataset Adequacy Scrutiny using Factor Exploration and Correlation. International Journal of Recent Technology and Engineering, Volume-8, Issue-1S4, pp 1105-1110.
13. Kalra, M., & Singh, S. (2015). A Review of Metaheuristic Scheduling Techniques in Cloud Computing. Egyptian Informatics Journal, 16(3), 275-295.
14. Kanaan, R., &Masa'deh, R. (2018). Increasing Citizen Engagement and Participation through eGovernment in Jordan. Modern Applied Science, 12(11), 225-231.