Introduction of Ontology based Query Retrieval Mechanism in Cloud Computing Architecture

Disha Grover	Barjesh Kochar, PhD	Y.S. Shishodia
Asst. Professor (IT)	Professor (IT)	Professor
Jagan Inst. of Mgt. Std	Jagan Inst. of Mgt. Std	Jagannath University

ABSTRACT

Cloud computing is one of the Information Communication Technology (ICT) trends. It is amalgamation of network of computers connected by centralized physical server. These servers are located on remote computers so that we can access applications directly from distant computer by sitting at our place. The paper makes readers aware of concepts of cloud computing along with its architecture. Few distinctive features among different cloud models have been described.

It is believed that existing model must be improved time to time to ensure efficient computing of tasks. Keeping this in mind, an improved architecture of cloud computing has been proposed in following paper. The proposed model consists of ontology module into cloud information architecture and retrieves terms related to user query using Query Retrieval Mechanism.

Keywords

Cloud computing, Network models, Security, Query Retrieval Mechanism.

1. INTRODUCTION

The phase of cloud computing began from telephone industry in era of 90's when virtual network service was introduced. The service reduces dependency on hardware among customers and suppliers. Although this virtualization service has advantages like uniform distribution of bandwidth and reliability but it suffered from following disadvantages.

 (i) It defines different paths for resource allocation due to changing time factors.

(ii) The servers connected to virtual network are not able to cooperate in non distributed environments.

It is given as name "dujour" says Gartner's Ben Pring [1]. Although cloud computing effect is ruling minds

of technologists, scientists and organizations but it is still occupied by some challenges as listed below:
Security [5] is one of the major element that needs to be addressed by cloud computing services during transmission of sensitive data and critical applications to cloud environments.

Table 1: Security issues associated with cloud computing

Layer	Components	Security Issues
End User or User layer	Cloud applications, programming tools and environment.	Security as a Service, Browser Security and Authentication
Service Provider Layer	Service Level Agreements (SLA) Monitoring, Scheduler and Dispatcher, Load Balancer, Policy Management.	Identity, Infrastructure, Privacy, Audit and Compliance, Cloud Integrity.
Virtualization layer	Number of Virtual Machines, Number of operating systems and its monitoring.	VM escape, Infrastructure, Identity and Access Management.
Data Center Layer	Servers, CPU's, Memory and Hardware.	Data Storage, Network and Server.

The paper is divided into following sections: Section 2 presents concise overview of cloud computing and its deployment models. Section 3 lists deficiencies in existing cloud computing open architecture and replaces it with proposed model. Section 4 describes graphical evaluation of importing and accessing ontology data from cloud using ANEKA platform. Section 5 concludes about given paper.

2. CLOUD COMPUTING AND ITS BENEFITS

The cloud is defined as resource that consists of routers, firewalls, gateways and storage servers. Cloud is called as Data Center enclosed with boundary called as Demilitarized Zone (DMZ). This zone stores confidential information. In simple words, we can say that cloud refers to Internet and its infrastructure.
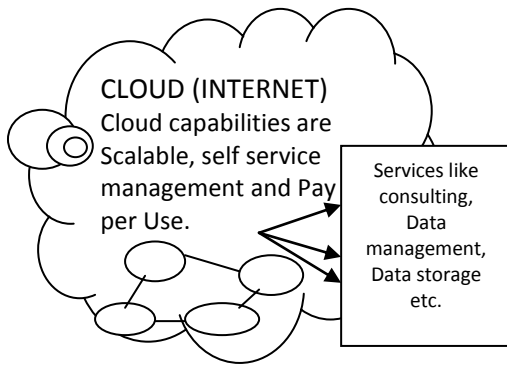
Fig 1: Capabilities of Cloud

Cloud computing services needs to be updated from time to time as per user perspectives and demands. Time to time updations in existing computer resources at various data centers is one of the factor that led to development of cloud computing. Doing updations led to advantage that it offers wide variety of services from software to security which are available all time and can be accessed from anywhere.

Use of cloud computing is useful to users because:

- It is inexpensive.

- It is convenient to use.

- Users can access data and use applications with the help of PC and Internet access.

- Software applications need not to be installed on computer; they can be directly accessible through Internet.

- Cloud computing produces applications to market very quickly by using most appropriate resources satisfying user needs.

Cloud computing performs services in ascending order i.e. we can use an acronym "ASC" which stands for Application, Storage and Connectivity.
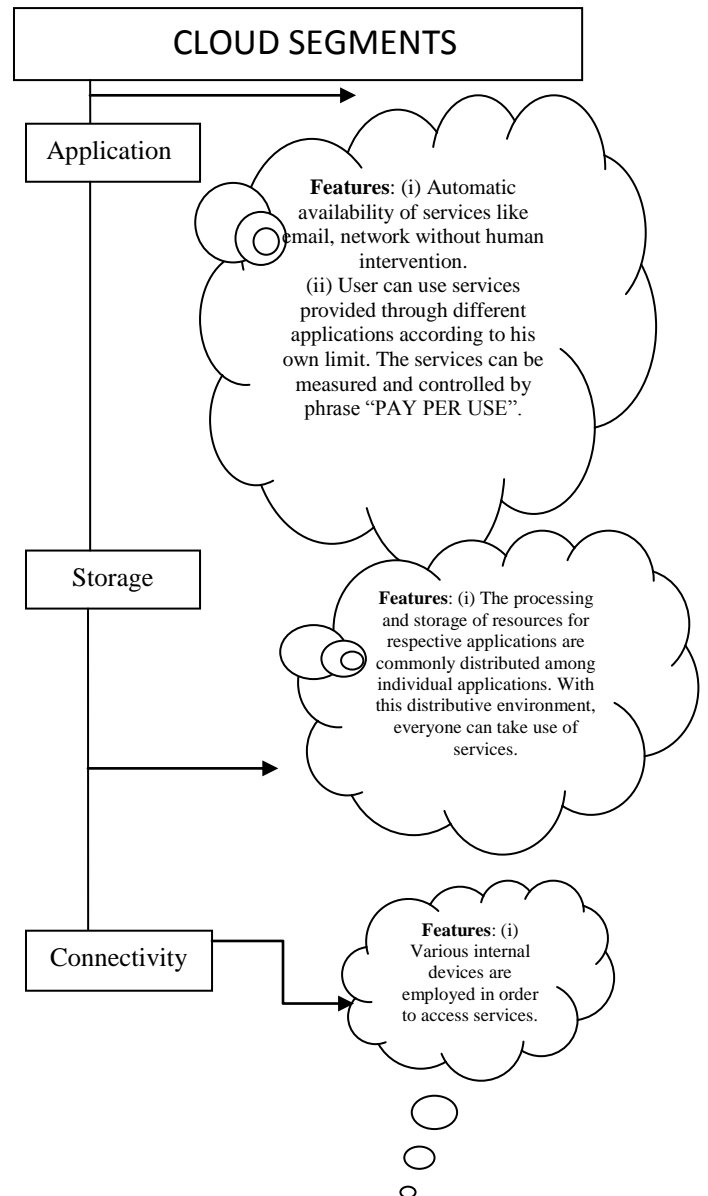
Fig 2: Features of Cloud Segments

The way in which cloud computing is different from other parallel computing types like grid computing, distributed computing is that roots of cloud computing are vast and spread across various organizations. Cloud computing is a practical approach to utilize resources, time and money in minimal way.

2.1 Deployment Models
They are also referred to as cloud computing types- Public Cloud, Private Cloud and Hybrid Cloud. Their distinction is shown in table.**.**

Table 2: Differences among deployment models

S.No.	Public Cloud	Private Cloud	Hybrid Cloud
1.	Simple and easy to use.	Monitoring is needed to control latest software updates	Most efficient (combination of both)
2.	Widely	limited	Used to

	accessible.	accessible	reduce work load.
3.	Less costly and reliable	More costly and less reliable	Most costly and most reliable.
4.	Suitable for handling large workload pressure	Not suitable for large workload pressure	Suitable for handling large workload pressure
5.	No space allocated for data center	Largest space allocated for data center.	Average space allocated for data center

3. CLOUD COMPUTING ARCHITECTURE

The need for cloud computing architecture is envisioned in order to produce following outcomes:

- Reduces problem of using multiple machines at one time.

- Maintain interoperability among different tasks on various machines.

- Context switching among users becomes easy in case of failure of one machine.

The layout of cloud computing is divided into modules wherein each module defines its associated services.

(A) **Cloud services module**- It consists of five abstraction layers:

a. *Hardware as a Service*: -This layer is termed as physical layer. It is bottom layer that comprises of cloud vendors, servers, operating systems, devices and switches. It performs data processing.

Use of cloud providers: - It provides information about hardware issues with help of remote scriptable boot loaders. These loaders allows cloud provider to specify the initial set of operations that are executed by servers during boot process.

b. *Software Kernel*: - This layer lies between hardware and software layer. This layer manages server's hardware resources and performs programs to run in parallel form.

*Implementation of S/W Kernel*: - It includes OS, Kernels, Hypervisors, Clustering and middleware.

Hypervisors: - It allows multiple operating systems to run on servers at same time.

Clustering middleware: - It is software located on group of servers which allows communication between multiple processes executing on different servers.

c. *Software Infrastructure*: - This layer is linked with application layer because it deals with new software applications that deliver services to users. Services in this layer are as follows:

(i) Data Storage as a Service (DaaS): - It stores data of users on servers located in remote locations. Users can access their information via web. Daas is evaluated on standards related to categories like Performance, scalability, relocatibility and

accessibility. Examples of clouds in DaaS is Amazon S3 (simple storage device).

Virtual machines (VMs)

Server
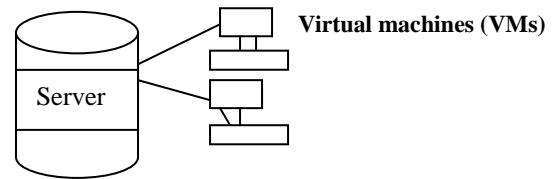
(ii) CaaS (communication as a service):- It provides communication that is reusable, schedulable, configurable and also encrypted. This communication enables CaaS to perform services like network security, real time adjustment to provide better bandwidth and network monitoring.

Examples include VOIP (Voice over internet telephone), E-mail and Video conferencing.

d. *PaaS (Platform as a service)*:- Users of this layer include cloud application developers who use applications to implement and distribute their resources via internet. Developers are provided with programming language and set of API's. It is also used for storage of resources.

Examples of S/W environments are GoogleApp engine and salesforce.com

e. *Application Layer or SaaS (Software as a service)*:- This layer acts as an interface between cloud applications and end users to offer them in demand. It is so because cloud users run programs by utilizing the computational power of servers and it reduces hardware requirements of machines.
In this layer, we don't have to install software on computers as all cloud software is located in providers' data centers.

(B) Cloud Infrastructure Management Module
This module comes under Infrastructure as a service layer that manages software and hardware with the help of virtual machines. They reduce workload on various cloud vendors while providing their services to users. It includes techniques like para-virtualization, Use of Virtual machines reduces the workload pressure on various cloud providers by following virtualization techniques like hardware virtualization, para-virtualization etc.

(C) Cloud Ontology generation module
This module transforms results by user into meaningful relationships called as Ontology. It retrieves information from collection of large documents and processes the results of search query. It is done by query retrieval mechanism.

(D) Cloud Authentication & Standards module

This module deals with factors like Quality of Service (QoS), monitoring cloud models and updating them according to latest technology trends, defines model standards according to Service Level Agreements (SLA's).

3.1. Other cloud computing architectures
defined in literature
(a) Soto mayor et al [2] proposed three layer models:

(i) Cloud Management: This layer provides remote and secure interfaces for creating, controlling and managing resources. It is available in IaaS.

(ii) Virtual Infrastructure (VI) management: VI management manages Virtual machines across multiple hosts.

(iii) Virtual Machine Manager: - It offers 3 facilities to Virtual machines on single host. Services are start, suspend and stop.

(b) Wang et al [3] proposed Cumulus architecture which has 3 layers:

(i) Virtual network domain

(ii) Physical network domain

(iii) Oracle File system.

3.2. Outline of proposed Cloud Computing Architecture

CLOUD SERVICES & INFRASTRUCTURE MODULE
(APPLICATION LAYER or SaaS- Includes pick and use services)

PLATFORM AS A SERVICE (PaaS- Includes data mining and user API's)

SOFTWARE INFRASTRUCTURE (IaaS- Includes DaaS and CaaS)

SOFTWARE KERNEL (hypervisors, middleware)

HARDWARE AS A SERVICE (HaaS- Includes switches, devices)

CLOUD ONTOLOGY [4] GENERATION MODULE
It processes search results of user and transforms them into hierarchical representation called as Ontology.

Database (Input) Preparation of Data Tabular Data Mapping of Data

Data Mining

Ontology Cloud (O/P) Ontology Builder Knowledge

Query Retrieval mechanism - It deals with retrieval of query terms from search results or derived ontology with the help of some transformation rules. At last, ranked results are available on cloud clients (web browsers, apps) that are accessed by users without any installation through SaaS. ACCESSING DATA FROM ONTOLOGY CLOUD CAN BE DONE USING ANEKA CLOUD PLATFORM. IT's GUI IS SHOWN IN NEXT SECTION.

CLOUD AUTHENTICATION & STANDARDS MODULE
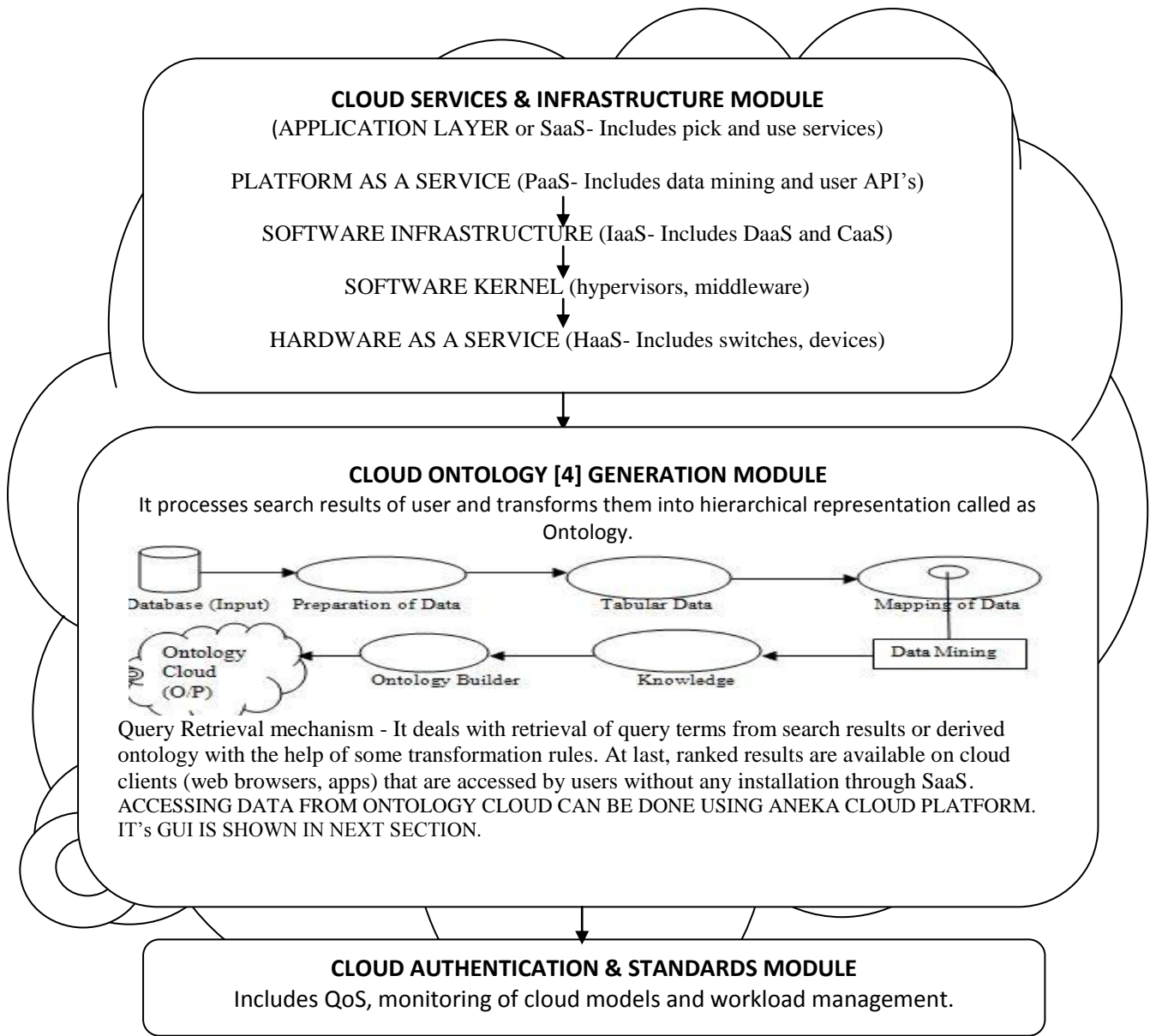Includes QoS, monitoring of cloud models and workload management.

Fig 3: An improved and compact model of Cloud Computing Information Architecture with enhancement of Query Retrieval Mechanism

3.3 Working of Query Retrieval Mechanism
The process of query retrieval is shown in following steps:

- Filters are used to distinguish between relevant and non relevant terms searched by user.

- Then filtered query process applies various ontology based rules to create a separate query which executes independently using full text search engine.

- After applying rules, we get transformed query corresponding to transformation.

- The query is executed and ranked results are produced on cloud clients (web browsers and applications).

- User can now access these results as per their cloud storage.

3.4 Anomalies found in existing cloud computing information architecture model

The existing cloud information architecture uses "GrepTheWeb" architecture for extracting documents from huge collection of online documents. But, it has few shortcomings:

- It requires two inputs before processing starts. One input is list of documents selected by user from Internet and second is Regular Expression. Regular expression is set of terms, phrases that are related to user's query.

- For maintaining two inputs, user has to devote lot of time by analyzing whole documents and it leads to more manual work.

- After user involvement, the application will respond. It retrieves subset of documents that matches given query.

- It does not filter documents selected by user and does not performs indexing on them. Indexing means creation of metadata.

- The existing model does not have provision of reusing the extracted terms for further need. Again user has to select documents from web and whole process repeats.

4. ACCESSING DATA FROM ONTOLOGY IN CLOUD

The cloud ontology generation module in proposed architecture involves retrieval of data from designed ontology in backend with the help of ANEKA cloud application platform. It requires adding dataset parameters and sharing of ontology file on cloud. Its GUI is shown below:
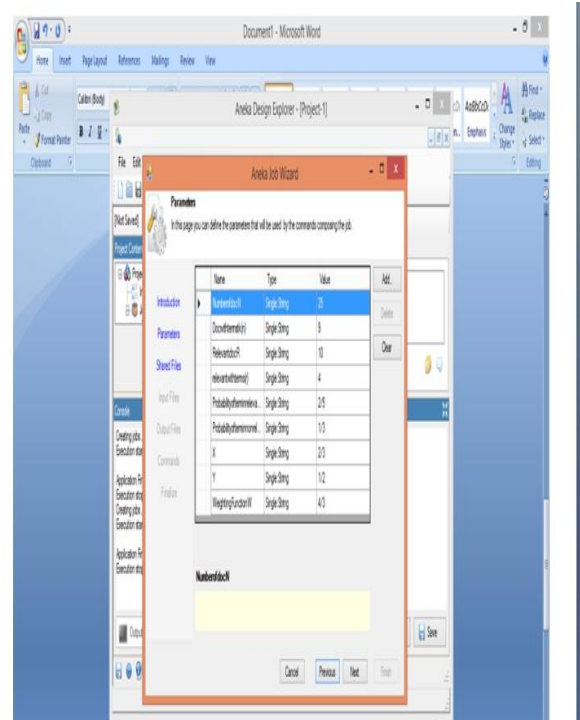

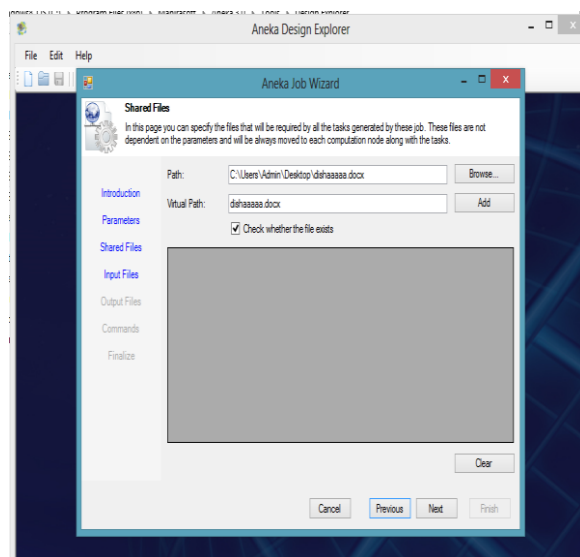**Fig 4: Adding given dataset parameters in ANEKA**

Fig 5: Given file ontology ccoa.docx is being shared by ANEKA platform

5. CONCLUSION AND FUTURE SCOPE

The paper presents pros and cons of cloud computing in world of information and technology. It describes cloud computing from infancy phase to developing phase. Its roots are still growing and occupying minds of various technologists and researchers. There are several models of cloud computing like Multi Source information architecture, security architecture, open architecture but none of them is able to provide any solution for refining of query terms extracted from web documents.

Its solution lies in concept of Ontology. It means to create relationships and associations between extracted documents so that user can easily identify those documents and use their services from cloud clients. Cloud providers provide these services as PAY PER USE. The proposed cloud information

architecture deals with ontology generation module for unified representation of resources and messages stored on cloud clients. After accumulation of data at one place in cloud, it becomes easier for user to access data and services using Query retrieval mechanism. This mechanism filters results and refines them as relevant or non relevant.

6. FUTURE SCOPE

Unlike Grep architecture, proposed model also provides provision of reusing produces results for future use. It makes our search refined and efficient because of ontology generation module. Further improvements can be made in this proposed model.

7. REFERENCES

[1] G. GRUMAN, E. KNORR, What cloud computing really means. *InfoWorld*, (2009, May). [Online].Available: http://www.infoworld.com/d/cloudcomputing/ what-cloud-computing-really means- 031

[2] SECURITY GUIDANCE FOR CRITICAL AREAS OF FOCUS IN CLOUD COMPUTING [Online].Available: http://www.cloudsecurityalliance.org/guidance/csaguide.pdf

[3] L. WANG, J. TAO, M. KUNZE, A. C. CASTELLANOS, D.KRAMER, W.KARL, Scientific Cloud Computing: Early Definition and Experience, In *HPCC '08*, pp. 825–830.

[4] Gagandeep Singh Narula, "Information Retrieval through Semantic Web-An Overview", In proceedings of CONFLUENCE-2012 at Amity School of Engg and Technology (ASET).

[5] V Krishna Reddy, Dr. L.S. Reddy, "Security Architecture of Cloud Computing", International Journal of Engineering Science and Technology (IJEST).

[6] "Security and high availability in cloud computing environments", IBM Global Technology Services Technical White Paper, IBM, June 2011

[7] Tim Mather, Subra Kumaraswamy, Shahed Latif "Cloud Security and Privacy", O'Reilly Media, 2009

[8] V. Krishna Reddy, B. Thirumal Rao, Dr. L.S.S. Reddy, P.Sai Kiran "Research Issues in Cloud Computing " Global Journal of Computer Science and Technology, Volume 11,Issue 11, July 2011

[9] Gartner's 2008 Data Center Conference Instant Polling Results: Virtualization Summary – March 2, 2009.

[10] Rao Mikkilineni, Vijay Sarathy "Cloud Computing and Lessons from the Past", Proceedings of IEEE WETICE 2009, First International Workshop on Collaboration & Cloud Computing, June 2009.