ISSN:0975-9646

Moving to the Media Cloud for Data and Content Integration Management

Zulfikhar Ahmad[1], Ashis Kumar Mishra[2],Asisha Kumar Jena[3]

[1,2]*Department of Computer Science & Engineering,*
[3]*Department of Computer Application,*
College of Engineering &Technology
Bhubaneswar-751003, Odisha, India

***Abstract*— The amount of information available to a person is growing day by day; hence retrieving the correct information in a timely manner plays a very important role. This paper talks about indexing document collections and fetching the right information with the help of Cloud computing platforms. This paper is an attempt at creating personal media cloud application for internet connected devices which talks about indexing and searching document collections and fetching the right information with the help of Media Cloud. This private Media Cloud enables users to store files and access the files from their internet connected devices at anywhere.**

Keywords- ***Cloud Computing, Media Cloud, Content Classification, Content Sharing***

I. INTRODUCTION

We are witnessing a media revolution. Not only the rich and varied types of media are being used, but also the enormous tide of media utilization is coming. To meet the great opportunities and challenges coming along with media revolution, the new technology and fundamental facilities with more powerful capability have become the most urgent demands. Simultaneously the adjustments of commercial model and industry strategy are automatically necessary to adapt to these changes. Fortunately, here comes the Cloud Computing just in time .Cloud Computing has emerged and advanced rapidly in the very recent years as a promising technology. Generally it can be seen as the integration of Software as a Service (SaaS), Platform as a Service (PaaS), Infrastructure as a Service (IaaS), and Hardware as a Service (HaaS) [3-4]. Cloud computing platform usually provides a shared pool of highly scalable, manageable and schedulable virtual/physical servers, storage, computing power, network bandwidth and so on with modest price. It has the greatest potential to provide a long term package solution for the media revolution if deliberately designed, deployed and integrated with the advanced technologies on media storing, processing and transmission, along with the rational commercial model and industry strategy.

Media cloud [1] or multimedia cloud [5] hereby presents when media revolution meets the rise of cloud computing. The emergence of media cloud not only has great impact on the related research and technologies such as the architecture of the cloud computing platform, media processing, storing, delivery, and sharing, but also has profound impact on the commercial model, industry strategy, and even the society. To deliberately consider all of these aspects and propose the overall solution is the ultimate goal but is beyond the scope of this paper. Here we discuss media cloud mainly focusing on the related research work and technologies. The objective of this paper is to introduce the concept of sharing the large amounts of media files or to provide a comfortable user interface for media files manipulation. This article intends to present Media cloud architecture, its distributed search engine and the content adaption modules. However, Media Cloud functionality is not limited to this description.

For the experimentation point of View, familiar operations for finding and sharing media with devices over the internet turn finally in an awful lot of clicking, searching, copying, and pasting. To cope with this problem, this paper describes a solution for bringing the cloud computing concept as middleware in which data will be stored and authorized user depending on its accessibility can view the data anywhere around the world. Today to access another system both person must be online, after this concept which I introduced in this paper it is not required both to be online. The solution describes a middleware that can be instantiated in Cloud computing called Media Cloud, for classifying, searching, and sharing media across the home domain and the cloud. The objective is to provide the maximum compatibility with the less effort from the user guide. For that reason, it aligns to cloud computing concepts so users no longer need expertise in or control over, the underlying technology.

My implementation has illustrated a framework of integrating the full-scale text indexing engine Lucene with the types of files stored in my system .This technique followed a general approach of keyword search followed by properties filtering in the Media cloud will very useful for sharing data with family, friends or colleagues. The solution describes the Folders of the respective types and the indexed folders are placed in the Media Cloud provided the platform by Salesforce free developers edition. The authorize end-users can access the data depend upon the files allowed to be access.

The remaining of the article is organized as follows:

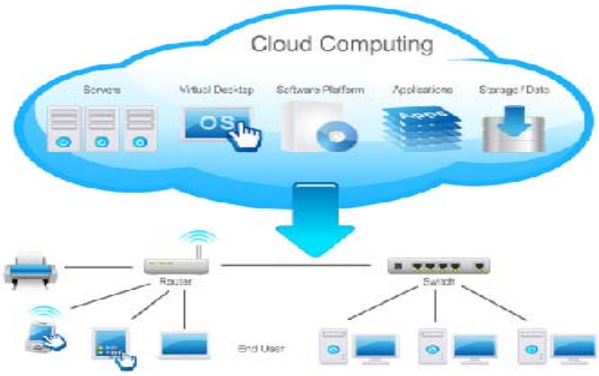
Section II briefly introduces the Cloud Computing paradigm and Media Cloud related to content management. Section III briefly gives the background foundation of this work and the related works in the field of document indexing and searching by Lucene. It also includes the SalesForce.com platform to transform our business into a Social Enterprises. In Section IV, we give the design of the proposed application and discussed about it interaction with the system. The implementation details

are explained in Section V together with some results for the deployment on the Media Cloud. Section VI summarizes the work with future work and conclusion.

II. Cloud Computing And Media Cloud

Cloud Computing is a technology that uses the internet and central remote servers to maintain data and applications. Cloud Computing allows consumers and businesses to use applications without installation and access their personal files at any computer with internet access. This technology allows for much more inefficient computing by centralizing storage, memory, processing and Cloud Computing is the delivery of computing as a service rather than a product, whereby shared resources, software, and information are provided to computers and other devices as a utility over a network. It is a model for delivering information technology services in which resources are retrieved from the internet through Web-based tools and applications, rather than a direct connection to a server. Data and software packages are stored in servers. However, Cloud Computing structure allows access to information as long as an electronic device has access to the web.

Cloud Computing is so named because the information being accessed is found in the "Clouds", and does not requires a user to be in a specific place to gain access to it. Companies may find that Cloud Computing allows them to reduce the cost of information management, since they are not required to own their own servers and can use capacity leased from third parties. Additionally, the Cloud-like Structure allows companies to upgrade software more quickly.



Cloud Computing as an Internet-based computing; where resources, software and information are provided to computers on-demand, like a public utility; is emerging as a platform for sharing resources like infrastructure, software and various applications. The majority of Cloud Computing infrastructure consists of reliable services delivered through data centres and built on servers. Cloud often appears as single points of access for all consumers' computing needs. Commercial offerings of the cloud are expected to meet quality of the services guarantees for consumer satisfaction and typically offer service level agreements. Cloud Computing means using multiple server computers via a digital network, as though they were one computer. Often, the services available are considered parts of Cloud Computing. Traditionally, without a cloud, a Web Server runs as a single computer or a group of privately owned computers. The computers are powerful enough to serve a given amount of requests per minute and can do with a certain amount of latency per request. If the computers' website or web application suddenly becomes more popular, and the amount of requests are far more than the web server can handle , the response time of the requested pages will be increased due to overloading. On the other hand, in times of low load much of the capacity will go unused.

Media content has become the major traffic of Internet and will keep on increasing rapidly. Various innovative media applications, services, devices have emerged and people tend to consume more media contents. We are meeting a media revolution. But media processing requires great capacity and capability of computing resources. Meanwhile Cloud Computing has emerged as a prosperous technology and the Cloud Computing platform has become a fundamental facility providing various services, great computing power, massive storage and bandwidth with modest cost. The integration of Cloud Computing and media processing is therefore a natural choice for both of them, and hence comes forth media cloud Consumers increasingly expect and demand content anywhere, anytime, and across a dazzling array of media and devices. To compete and succeed in this environment, organizations of all kinds need more efficient, cost-effective ways to create, store, manage, share, and deliver digitized content. For many, Cloud capabilities tuned specifically for media and entertainment requirements may be the solution. The objective of this solution is to help users to create a community for managing contents located inside or outside their local networks.

III. Background

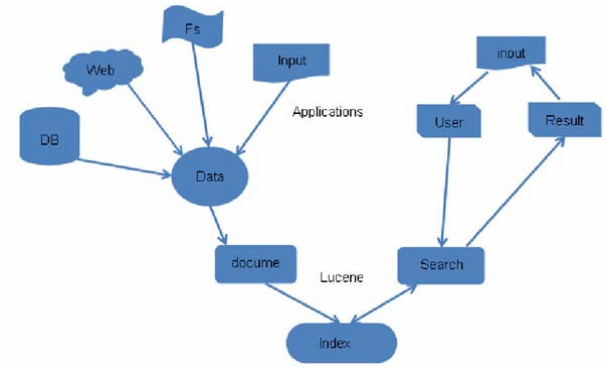
The heart of all search engines is the concept of Indexing; Indexing in short can be defined as the processing of original data into a highly efficient cross reference lookup in order to facilitate rapid searching. Indexing can be defined in a much better manner by taking an example. Let us assume that a person wants to search for a word or a phrase among a large number of files. The simplest method would be to sequentially scan each file for the given word or phrase. The main disadvantage with this approach is that it does not scale to larger file sets or cases where files are large; hence to search large amounts of text quickly, one must first index that text and convert it into a format that will let the person search it rapidly, eliminating the slow sequential scanning process. This conversion process is called indexing, and its output is called an index.

A Lucene index is stored in a single directory in the file system on a hard disk. The core elements of a Lucene index are segments, documents, fields, and terms. In a Lucene index every index consists of one or more segments. Each segment contains one or more documents. Each document has one or more fields, and each field contains one or more terms. Each term is a pair of Strings representing a field name and a value. A segment consists of a series of files. Pictorially the entire structure of a Lucene index can be represented as shown in figure below.

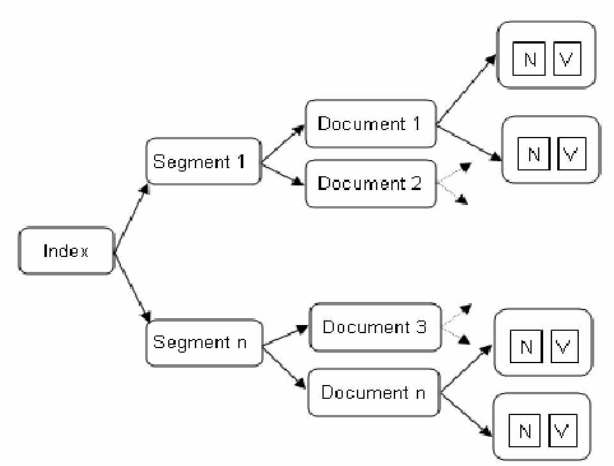
To optimize an index, one has to call optimize () on an IndexWriter instance. When optimize () is called, all in-memory

documents are flushed to the disk and all index segments are merged into a single segment, reducing the number of files that make up the index. However, optimizing an index does not help improve indexing performance. As a matter of fact, optimizing an index during the indexing process will only slow things down. Despite this, optimizing may sometimes be necessary in order to keep the number of open files under control. For instance, optimizing an index during the indexing process may be needed in situations where searching and indexing happen concurrently, since both processes keep their own set of open files. In fact, if more documents are added to the index, one should avoid calling optimize (). If, on the other hand, one should know that the index will not be modified for a while, and the index will only be searched, and it should be optimized. That will reduce the number of segments, and consequently improve search performance, the fewer files Lucene has to open while searching, the faster is the search.

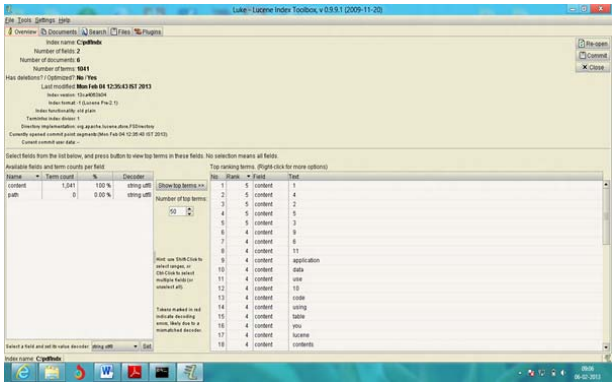
It is possible to add indexing and searching capabilities to any application using Lucene. It is possible to index and make searchable any data that can be converted to a text format. Lucene is independent of the source of data, its format, and even its language as long as it can be converted to text. This means one can use Lucene to index and search data stored in files, web pages, on remote web servers, documents stored in local file systems, simple text files, Microsoft Word documents, HTML or PDF files, or any other format from which one can extract textual information. Similarly, with Lucene one can index data stored in databases, giving users full-text search capabilities that are absent in many databases. The basic architecture of a Lucene application can be visualized as shown in figure 2.2



A Lucene index is an inverted index. An inverted index means that the content of the documents that are analyzed has their important terms indexed as a pair consisting of a field name and a field value. A field contains many terms that point to the corresponding documents. To summarize, an inverted index makes the process of retrieving documents from a system a breeze. Finally, the documents are searched in the fields and in their values. A Lucene index consists of many segments. A segment is created every time when new documents are created and indexed. Hence each segment has many documents stored in it. The documents consist of indexed Fields. An indexed field is a pair consisting of a field name and a field value pairs. Fields are used for calculating weights and ranking search results.



An actual Lucene index screen capture is shown in figure2.4, it is the index of a PDF document, as it can be seen in the figure it consists of three fields and it is the index of only one PDF document.

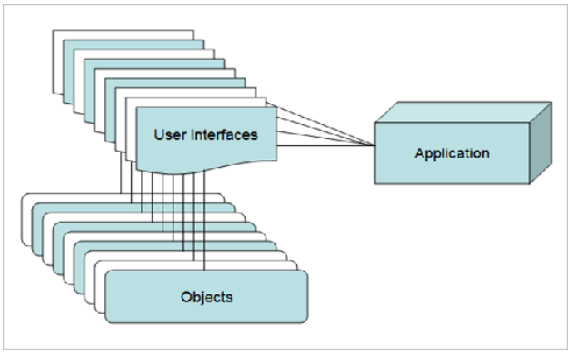


Building and running business applications with traditional software has always been too complex, slow, and expensive. A new model, called cloud computing, has emerged over the last decade to address this problem. Applications that run in the cloud are delivered as a service so companies no longer have to buy and maintain hardware and software to run them.

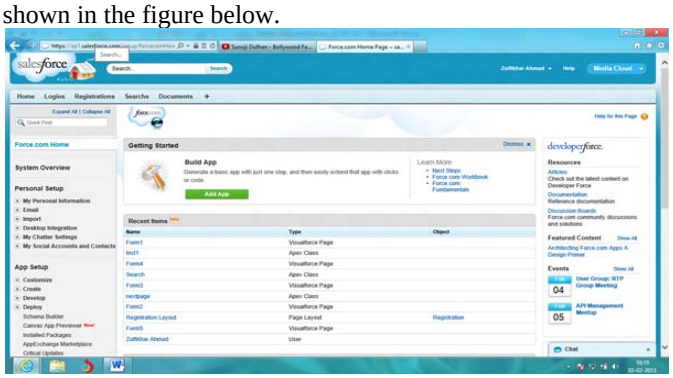
Salesforce.com pioneered this model with our applications business over the last decade. More recently, we have opened up our infrastructure and made it available for anyone building any business application. These business applications run on our servers using the Force.com platform.

The Force.com platform allows us to store structured data, implement business logic with workflow rules, approval processes and custom code, support Web browsers and mobile devices, integrate with other applications, do reporting and analytics and scale up or down-all with sub second response time, high availability, and security we need to run our mission critical business apps. The Force.com platform is the fastest path to building complex enterprises apps. Unlike a stack of separately designed hardware and software products, the Force.com platform speeds innovation through a powerful yet easy-to-use development and deployment model. Users easily assemble applications, then instantly deploy them to the salesforce.com infrastructure.

The Force.com platform consists of a few core components. These core components define the outline of all applications you will build on the platform.
The figure below shows the way that these core pieces relate to each other.



Cloud computing, based on the amount of recent coverage, and may seem like a brand new development, but Salesforce.com has been delivering applications from the cloud for almost a decade. The Force.com platform supports almost 50,000 customers, containing more than one million individual subscribers, running enterprise applications offered by Salesforce.com, its associated independent software vendors (ISVs), and in-house and independent developers. There are more than 70,000 custom applications running on the Force.com platform at the time of this writing, and the platform supports more than 150 million transactions a day.
With the Force.com Platform, we get a complete a stack-including a complete user interface to our data, full reporting and analytic capabilities, a flexible security and sharing model, and other services not displayed in the diagram, such as workflow and approvals-fully available and integrated from the outset. We can even leverage extended features such as built-in internationalization, full support for mobile devices, and integration with existing systems.
We access the Force.com platform through a Web browser, as shown in the figure below.



IV. Proposed Architecture

Media Cloud is the bridge to an open architecture that allows users to join their data to constitute a Cloud.
Media Cloud abstracts the underlying complexity to provide a new content distribution model that simplifies classifying, searching and accessing user-generated and commercial content within the network.

Media Cloud pursues fulfilling two goals:

- **Content Classification**
- **Content Sharing**

ISSUE:

An average user generates contents very quickly and stores them away in several devices. In fact, users' media library stops growing only since they reach storage capacity until they buy an additional hard drive or computer.

Problem:

Thus, it is usual for user to expand big time trying to find contents that were previously stored in his/her own devices, organize them into collections, and manipulate them to produce new formats or presentations.

Solution:

The first goal i.e. Content classification alleviates the above problem by providing an indexing service for searching the large set of contents.

ISSUE:

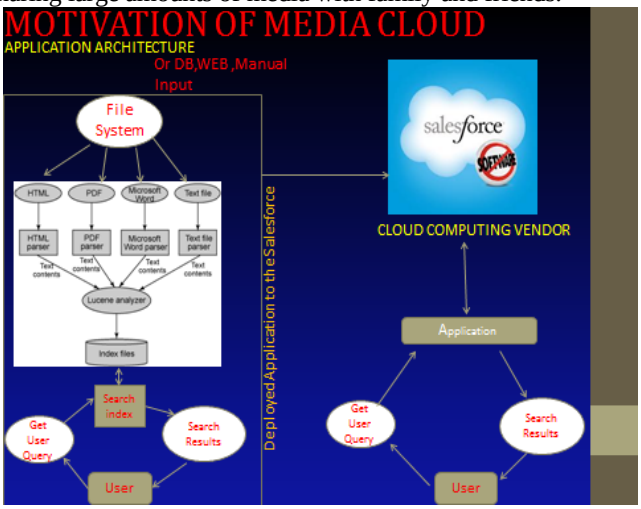
Consider a group of friends that have just arrived from a half month travel and they all want to exchange the pictures stored in their cameras.

Problem:

An option would be to upload every picture and video to a social network to let others to individually select and download the pictures they like. However, viewing, selecting and downloading individual media file from the social network would be pretty tough and time consuming process.

Solution:

The second goal i.e. Content Sharing alleviates the problem of sharing large amounts of media with family and friends.



V. Implementation

The Implementation took place in two phases:

- ✓ Content Classification
- ✓ Content Sharing

Content Classification

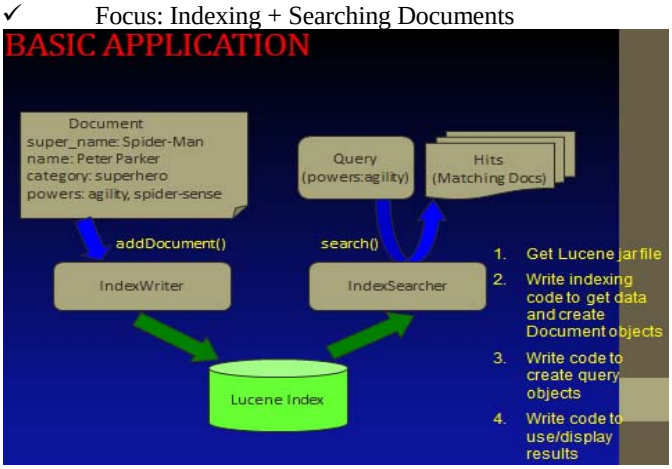
In 1970s, information stayed only in certain places with certain people in a certain format. In 2010 era, information is at your finger tip due to the huge revolution in the IT industry. When you have tons of details, search plays a vital role to extract the right information at the right time.

Software Specification:
Eclipse IDE
Tomcat server
Lucene tool
What is Lucene?

✓ Apache Lucene a high-performance, full-featured text search engine. It's written in Java as open source.

✓ High performance, scalable, full-text search library

✓ Focus: Indexing + Searching Documents

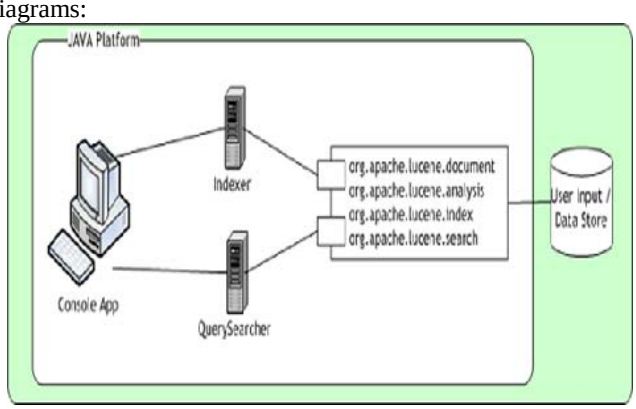


Then please think about a common problem:

How can I find in which retrieved documents that include "Department of Computer Science"?

At the core of Lucene's logical architecture is the idea of a **document** containing **fields** of text. This flexibility allows Lucene's API to be independent of the file format. Text from PDFs, HTML, MSWORD & Open document documents as well as many others can all be indexed as long as their textual information can be extracted.

The Content Classification goal is fulfilled by the following diagrams:



CONTENT SHARING GOAL

Salesforce.com is a software-as-a-service Company, which is now also offering platform-as-a service. Salesforce is a pioneer in providing a high quality on-demand CRM software (referred to as " The Sales cloud " by Salesforce).They have diversified into Customer service and support on-demand Offerings as well (Referred to as "The Service cloud by Salesforce).Salesforce is a classic manifestation of two (SaaS and PaaS) out of the three major components Of Cloud computing.

Then please think about a common problem:

How can I want to share my files to my family and friend even if I am not in the home domain and they can access my files from anywhere in the world?

The complete application of content classification is deployed in the cloud computing platform. Then build Media Cloud application in the SalesForce.com by login into free developer's edition namely developersforce.com .Build an interface for login, registration, search and documents for accessing the files as per the authorization to access it.

VI. Future Work And Conclusion

During the study of recent research works on media cloud, we find that there might be increasing disconnection between research on media cloud and business models and industrial deployment. There are many media cloud system already deployed in practice. For example, although many details of Google TV [14], Apple TV[15], Hulu[11], Youtube[76], Netflix[17], and Last.FM[12], and PPLive[18], are not revealed, we can conjecture that they are evolving and moving to the media cloud. But the related research work seems to lag behind the development in industry. This could be at least partially explained by the following statements: "The challenge now is to get ahead of that change, and to meet the consumption requirements of the future."[1], and "In the coming content-driven world, smart companies are moving their media to the cloud." [1]. So this time the industry companies seem to be faster than our researchers.

Conclusion

Content-dependent organizations of all kinds are struggling to produce, store, share, and distribute massive new volumes of digitized material. They serve consumers who now expect content to be readily available at any time, in any place, and across a growing multitude of devices and media. Forward-looking organizations are now using advanced media cloud capabilities to better manage content across the value chain. The power and flexibility of cloud computing support sector-specific solutions for telecom service providers, cable network operators, and media and entertainment firms—and for virtually any company or agency that uses or shares digital content. A media cloud can be deployed to reduce both capital and operating costs, and to extend the useful life of in-place infrastructure. The cloud can be used to reach new market segments, to drive long-term revenue, and to introduce productive new advertising strategies. Organizations can use cloud-based capabilities to enhance customer service, satisfaction, and spend. In the coming content-driven world, smart companies are moving their media to the cloud.

Media Cloud provides an easy to manage, cost-effective solution for bringing Cloud Computing paradigm to content sharing among federated networks. The cost-effectiveness is achieved by sharing resources that could be underused in other cases. Media Cloud encourages cooperation among networks facilitating media classification, management and sharing. Distributed search and content delivery over the cloud are among the most important features of Media Cloud.

REFERENCES

[1] S.Poehlein, V. Saxena, G.T.Willis, J.Fedders, and M.Guttman. (2010, 8/15).MovingtotheMediaCloud.Available:http://h20195.www2.hp.com/V2/GetPDF.aspx/4 AA2-1000ENW.pdf

[2] Cisco-White-Paper," Cisco Visual Networking Index- Forecast and Methodology, 2010-2015," June 1, 2011.

[3] M.Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G.Lee, D. A Patterson, A. Rabkin, I.Stoica, and M.Zaharia, "Above the Clouds: A Berkeley view of Cloud Computing," EECS Department, University of California, Berkeley UCB/EECS-2009-28, February 10 2009.

[4]N.Carr(2006)HerecomesHaaS.Availablehttp://www.roughtype.com/archives/2006/03 here_comes_haas.php.

[5] Z.Wenwu, L.Chong, W.Jianfeng, and L.Shipeng, "Multimedia Cloud Computing," Signal Processing Magazine,IEEE,vol.28,pp.56-59,2011.

[6] B. D. Johnson, Screen Future: The Future of Entertainment, Computing and the Devices We Love: Intel Press, 2011.

[7] CNN. CNN website. Available: www.cnn.com

[8] NBC. NBC website. Available: http://www.nbc.com/

[9] Google. Google website. Available: http:// www.google.com

[10] Facebook. Facebook website. Available: www.facebook.com

[11] Hulu. Hulu website. Available: www.hulu.com

[12] Last.fm. Last.fm website. Available: http://www.last.fm

[13] K. kalevi, "Quality of Experience in Communications Ecosystem," Journal of Universal Computer Science, vol. 14, p.9, 03/01/2008 2008.

[14] Google Google TV website.Available:http://www.google.com/tv/

[15] Apple Apple TV website. Available:http://www.apple.com/appletv/

[16] YouTube. YouTube website. Available:http://www.youtube.com/

[17] Netflix. Netflix website: Available: http://www. Netflix.com/

[18] PPLive. Available:http://www.pptv.com