Data-warehousing on Cloud Computing

Hemlata Verma(Assistant Professor-IT)
Department Name : Information Technology
Name of Organization : Tecnia Institute of Advanced Studies
City : Uttam Nagar, Country : India

Abstract[1][2]: Our everyday data processing activities create massive amounts of data.Cloud Computing has emerged as a new paradigm for hosting and delivering services over the internet. Cloud computing is attractive to business owners as it eliminates the requirement for users to plan ahead for provisioning and allows enterprises to start from the small and increase resources only when there is a rise in service demand. In this paper our issue is data warehouse on cloud computing, the main aim of this paper is how to create a datawarehousing over cloud computing, recycle & reuse data, recover the data and updation of the information through datawarehousing over cloud computing.

Index Terms: Datawarehousing, Cloud Computing, Data, Algorithms, Technology

Introduction[3][4]: With the rapid development of processing and storage technologies and the success of the internet. Cloud computing is a model for enabling convenient, on demand network access to a shared pool of configurable computing resources such as: networks, servers, storage, applications and services those can be rapidly provisioned and released with minimal management effort or service provider interaction. Cloud computing is basically for storing and accessing of applications from the computer(Remote).

Whereas Datawarehousing**[5]** refers to the combination of many different databases across an entire enterprise used to store the information and generate the query regarding the required data. The sources will help to access the information, save downloads & update the information viz. suppose your one file has some data and you want to add same data or updating the data in a file, so these sources will help you. We want to see the data so we can access the data or generate a query for getting the information. Our concern is how to manage the data through recycle, reuse, reduce & recover information in cloud environment. The responsible usage of the cloud should take place to provide a better environment for all.

Managing Stored Data[5]

How do we deal with stored data? We need effective strategies to handle the stored Datawarehouse data. To provide guideline for stored data management we turned to the techniques used in real life. In this paper we will discuss about how to manage

411
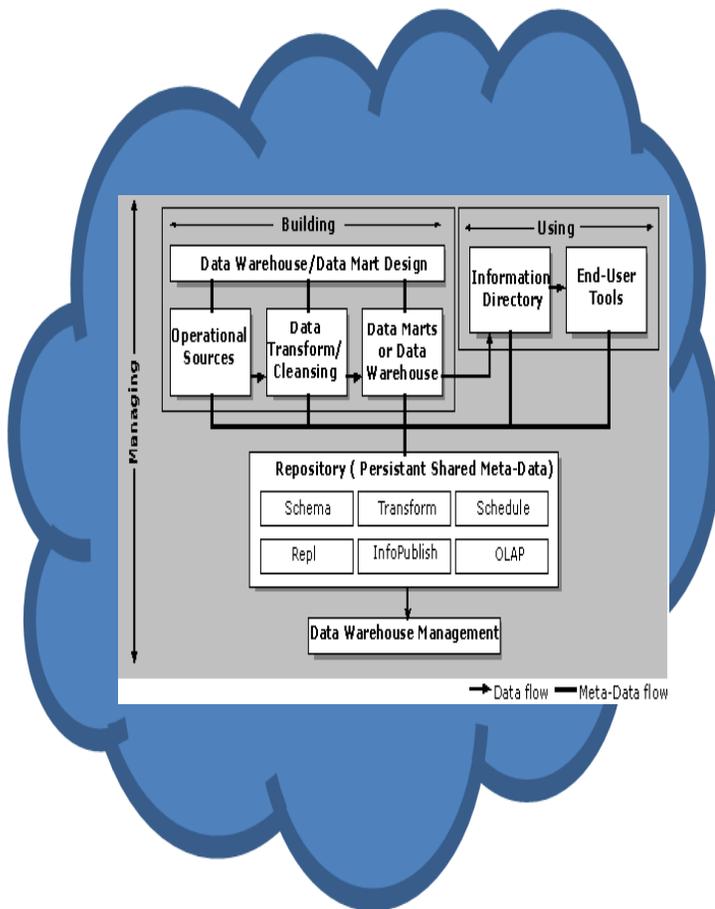
Fig. 1.1 Datawarehousing on Cloud Computing[6]

Datawarehouse with cloud computing . Data warehousing over Cloud computing has recently emerged as a compelling paradigm for managing and delivering services over the Internet. In a datawarehouse all data from an organization can be brought together in one place. Data warehouse systems require a different kind of database. Data warehousing systems are also used for complex analytics involving huge amounts of data (OLAP or online analytical processing). Data warehouse may support OLAP(on line Analyatical processing) tools, allowing the decision maker to navigate the data in the data warehouse.

The following steps to create Data Warehouse over cloud Computing. **[6][7]**

1. Data is stored in various data sources.
2. Develop the tools which are used for data extraction from the various sources.
3. After extracting the data then transform the data in various dimensions.
4. The data is loaded into the data warehouse.
5. Ready for usage by OLAP and data mining tools.
6. Analyze the data

7. In addition to one data warehouse where all data come together, an organization may also choose to use data marts which carry only past data of the data warehouse. Data marts are more specialized and therefore easier to deploy.
8. A data mart is set up by a single department or division within an organization for a single purpose.
9. Then quickly implement a needed system, without affecting or changing the existing data warehouse. We move towards the cloud is just as relevant for data marts as for data warehouses.

The fundamental requirements of a data mart are simpler and human resources to be managed on the department level.

For Example: WallMart

The Data warehousing Methodology is organized into the following phases : **[8]**

Initation : Evaluating Readiness and Opportunities

Analysis : Analysis and Requirements Determination

Design : Datawarehouse and Data Mart Models(Star Schema/ Multidimensional Model),Technical Architecture, Obtain Datawarehouse Inputs

Construct: Data Load and Presentation/Analysis Tools

QA : Test

Rollout : Deploy in Production

Iterate : Make Incremental Changes

In the following ways we can manage the data :

Datawarehousing over Cloud Computing

1. **Reduce[9][10]   :** We use compression techniques to store the data. Compressed data will help to save space and we can store large data. A good example is "zipping" up documents to send in emails. … There are two main types of compression algorithm: lossless and lossy.

Lossless algorithms, as the name suggests, convert the data without any loss of the original file whatsoever. It compresses the amount of data that the file takes up, but we will can then uncompress the file and it returns back to the exact original file. For lossless compression, good examples of where we use this technology are spreadsheets and text files where we need the data.

Lossy algorithms are more efficient than lossless algorithms, but if you uncompress them it's impossible to get back to the … original file. Lossy compression can be used in scenarios such as video streaming and photographs, where what [generally happens] is you get slight loss of quality of the file but it's not discernable to the naked eye.

Deduplication by contrast reduces the size of large data sets by removing information that's duplicated and leaving a pointer to the original data. So, we have got one copy of the data and pointers to that [from where other examples of the same] data used to be.

Data deduplication can work at the file or block level. As an example, if I were to send an email to 20 people with an attachment, there'd be 20 copies of the attachment in the email system. A data deduplication appliance could do & will keep the original copy of that attachment and then put pointers to it from where the other copies of the file would be. So, data deduplication can be very efficient where you have lots of copies of user file data or lots of pages of data that are the same.

Deduplication is being used for a long time in backup applications but is becoming a more acceptable method for saving space in primary file systems as well.

413

2. **Reuse [11]:** Some data is already stored and that data is required for use.

Data can be reused and we can do some modifications on that data.

3. **Recycle [11]:** Deleted data is stored in Recyclebin with the help of Recycle bin we can extract the usable components and data.

4. **Recover [11]:** Deleted data can be recovered from various softwares and recycle bin & with the help of recovered data we can manage the data.

Advantages of Data Warehousing over Cloud Computing[12]

• Cost Reduction – Cloud Computing can reduce the paperwork, transaction cost, hardware cost and IT staff.
• Scalable – like electricity, water and pay-as-you go phones, some cloud computing services are billed, based on the amount of usage. Therefore, you only pay for what you really use and can easily upgrade your service without having to make costly additions to hardware or software.
• Right level for the right size of business – cloud-computing services are available in small and mid-sizes. This will reduce cost on software licenses like remote control software as well as server cost.
• Easier to collaborate – with cloud computing PC Remote Access is also possible. Meaning, users can access anywhere, anytime, thus can be collaborated with remote employees.

Conclusion

Data warehousing over the cloud computing[13] has potential for elasticity, scalability, deployment time, reliability and reduced costs. The capabilities of the Data warehousing over cloud computing is high, parallel and distributed. High Security issues will involved in the decision of moving a data from Data warehousing or data marts into the cloud.

It involve easier to control the environment. Datawarehousing over cloud computing provides various facilities to customers like self service provisioning, self service data management, web based upload and download the data and services are delivered over the network. IT also provides abstraction between hardware and computing software.

Data warehousing over the cloud is largely hypothetical and for example[14] benchmarking the cloud may lead to new insights in the possibilities and impossibilities of deploying data warehousing into the current-date cloud environment. Support by Amazon for IBM's DB2 and Oracle's database 11g is promising and testing these recently added capabilities may lead to more insights in the applicability of these systems in the cloud. The phone company requires regulatory approval to raise your rates; but that is not the case with Amazon or Google or Layered Technology. Granted that strong incentives exist to exploit the network effects (economies of scale and Moore's Law like pricing). It is a familiar and proven revenue model to give away the razor and charge a little bit extra for the razor blade. Technology lock-in! It is an easy prediction to make that something like that will occur once the computing model has been demonstrated to be scalable, reliable and popular.

414

References :

[1] Journal of Internet Services and Applications, "Cloud computing: state-of-the-art and research challenges", Page. No 1

[2] Vishal Jain1 and Mahesh Kumar Madan 2, "Information retrieval through Multi – Agent System with Data Mining in Cloud Computing", "International Journal of Computer Technology and Applications, Volume 2 :Issue 4, Page 1,2012

[3] Veena Goswami (KIIT University, India), Sudhansu Shekhar Patra (KIIT University, India), and G. B. Mund (KIIT University, India), "Optimal Management of Cloud Centers with Different Arrival Modes for Cloud Computing Environment","**International Journal of Cloud Applications and Computing",** Volume 2 :Issue 3, (pages 86-97), 2012

[4] Vaibhav C.Gandhi, Jignesh A.Prajapati and Pinesh A.Darji,Parul Institute of Engineering & Technology,Gujarat Technological University, "Cloud Computing with Data Warehousing", International Journal of Emerging Trends & Technology in Computer Science, Volume 1: Issue 3, Page 72,73, September – October 2012, ISSN 2278-6856

[5] William H. Inmon, Book, "Building the Data Warehouse". Page. 1-30

[6] A Research Paper on Data Warehousing,from http://www.oocities.org/dwarepk/#References, Nauman Mazhar, Ammar Sohail, M. Arshad Mughal, Aqsa Khursheed, Aqil Bajwa , Page. No 10

[7] Chantal Reynaud,Université Paris-Sud, CNRS (LRI) & INRIA (Saclay – Île-de-France), France,Nathalie Pernelle,Université Paris-Sud, CNRS (LRI) & INRIA(Saclay – Île-de-France), France, Marie-Christine Rousset,LIG – Laboratoire d'Informatique de Grenoble, France,Brigitte Safar, Université Paris-Sud, CNRS (LRI) & INRIA (Saclay – Île-de-France), France,Fatiha Saïs,Université Paris-Sud, CNRS (LRI) & INRIA (Saclay – Île-de-France), France, "Data Extraction, Transformation and Integration Guided by an Ontology", "International Journal of Data Warehousing Design and Advanced Engineering Applications: Methods for Complex Construction", Volume 1 :Issue 6, Page 18, 2012

[8]www.infogoal.com, http://www.infogoal.com/datawarehousing/data_warehousing_and_business_intelligence_methodology.htm

[9]http://www.computerweekly.com/podcast/Data-reduction-techniques-for-primary-storage-How-suited-are-deduplication-and-compression

[10] International Journal of Cloud Computing and Services Science, "Towards Information Security Metrics Framework for Cloud Computing", Muhammad Imran Tariq,Department of Computer Science and Information Technology, University of Lahore, Pakistan, Page. No. 209,210.

[11] Ragib Hasan, "The Life and Death of Unwanted Bits": Towards Proactive Waste Data Management in Digital Ecosystems,Randal Burns Department of Computer Science,Pg. No 3 & 4 Johns Hopkins University 3400 N. Charles Street, Baltimore, MD 21218, United States

[12]**http://www.articlesbase.com/software -articles/cloud-computing-pro- and-cons-5629637.html**

[13] Kees Van Gelder,"Elastic Data warehousing in the cloud is the sky really the unity", Faculty of exact sciences Vrije University, Amsterdam, the Netherlands Page. No. 5,15,16,17

[14] Amazon Web Services. Amazon Elastic Compute Cloud homepage. http://aws.amazon.com/ec2, November 2008.