

The Intelligent Decision Supporting Technology of Cloud-Based Public Data Warehouse

Chunsheng Zhang

College of Computer Science and Technology, Inner Mongolia University for Nationalities

Tongliao 028043, Inner Mongolia, China

Tel: 86-475-831-3358 E-mail: zhangcs_817@163.com

Received: December 18, 2011

Accepted: December 28, 2011

Published: March 1, 2012

doi:10.5539/cis.v5n2p75

URL: <http://dx.doi.org/10.5539/cis.v5n2p75>

Abstract

According to the current development of cloud computing and limitations of data warehousing and data mining technology, we proposed The intelligent decision supporting technology of cloud-based public data warehouse and gave the principle of the technology, the implementation process, and constraints the key issues of development prospects. The study of this technology can be realized solution of knowledge discovery in many fields of research within the global. The knowledge it produces can not realized by any other technical so far. It will be direction of development of data mining technology in the future.

Keywords: Cloud computing, Data warehouse, Data mining, Decision supporting

1. Introduction

Data mining technology has gone through five stages: the **first** generation is **a separate algorithm, a single system, a single machine, using the vector data**. The **second** generation is **combined with the database, support for multiple algorithms**. The **third** generation is **integrated with predictive models** to support Web data, semi-structured data, is a network computing. The **fourth** generation of distributed data mining, grid computing is based on **a variety of algorithms**, distributed across multiple nodes on the way. the **fifth**-generation cloud-based data mining and parallel modes of service, the same algorithm can be distributed across multiple nodes in parallel between multiple algorithms, multiple nodes in the implementation of on-demand computing resources, and distributed computing model using cloud computing model, data using DFS or HBASE, programming mode with a Map / reduce in this way (Ni xianjun, 2007; Mladen A. Vouk, 2008; LuisM Vaquero, 2008; Wang Lizhe, 2008; Francesco Maria Aymerich, 2008; Gianni Fenu, 2009).

Based on Cloud-data mining began China Mobile's largest cloud computing platform in China. At the end of 2008 Institute of Computing Technology Chinese Academy of Sciences and China Mobile completed the development of cloud-based data mining software PDM iner, the software integrates a variety of algorithms it can effectively solve a variety of cloud computing problems (Zhang Jianxun, 2010; Kuang Shenghui, 2010; Wang Peng, 2009; Wang Jiajuan, 2010; Xiang Guibing, 2010).

As the cloud of mass data storage and distributed computing, cloud computing environment for the massive data mining provides new methods and tools to effectively address the massive data mining of distributed storage and efficient computational problems.

Based on previous data mining data warehouse is often a certain unit or a department, with the rise of cloud technology, the techonogy of data mining face challenges.

Because of data limitations the current global data warehouse and knowledge discovery is almost impossible, cloud computing technology has brought way to this issue, through cloud computing technology to collect large amounts of data around the globe, and construct a large-scale data warehouse, based on these, data mining technology can get the rules than ever befor, consequently, it can realize disease diagnosis, natural disaster prediction and many other research work, and knowledge sharing for all mankind.

In this paper, we combine the data warehouse, data mining, cloud computing technology propose intelligent decision supporting technology of cloud-based public data warehouse that can build a data warehouse in the cloud model, collect data automatically within the world, through data mining algorithms to be constantly

updated rules for global use.

2. Cloud-based Data Warehouse to Build Public Works

Cloud-based storage cloud on the public database of user self-complete data collection, to become effective after pretreatment data warehouse.

Definition 2.1 Fixed users: Hire cloud computing center and use the appropriate data warehouse users as fixed users.

Definition 2.2 occasional users: do not hire the cloud data center storage, but for the support of public utilities, data voluntarily provided by users of the data warehouse.

Definition 2.3 Warehouse Model: The standard system-defined data warehouse data model, including the structure and relationships between tables.

Definition 2.4 free model: the occasional casual user with the characteristics of the user's own data warehouse model.

Definition 2.5 The original definition of a data warehouse: data warehouse system definition, the original data is stored.

Definition 2.6 effective data warehouse: data mining after pretreatment treated, can be directly used for data mining data warehouse.

The basic principles of building public database of the warehouse based on cloud:

In process of using data warehouse system by fixed users it produced a large amount of data which is stored in the original data warehouse system.

In order to support for public issues occasional users provides a lot of raw data according to their own data warehouse mode, this part of the data structure and data type may be different from the system data warehouse model, in particular, there may be a certain number of false data, in order to use these data, it must converse the data indeed.

The free data Occasional user produced must be noise-filtered first (the authenticity of identification data, to remove false data and abnormal data), to format converse according to the system standard model format n, including the deletion of attributes, properties, format conversion and so on. Conversion losses in the original data warehouse.

The occasional user-supplied data into the data warehouse data efficiently is the key to building a data warehouse and it is also challenging issues here, including:

(1) Determine the authenticity and validity of the data (noise filter) by manual or automated means remove the data which is incorrect and unreal.

(2) Converse the data which is reliable of the occasional user into the data warehouse effective data.

There are some differences between data provided by occasional users and data from effective data warehouse, first make properties correspond by distribution, including correspondence of meaning and format, that is, the meaning and properties must be integrated. Second, remove the invalid attributes and effectively predict empty properties of data warehouse.

(3) Add logo

Uniquely identifies is the basic conditions for physical integrity of the data warehouse. According to the identity of the effective data warehouse it is possible to the automatically generated identifies without duplication, which are wanted.

The data of original data warehouse can not be directly used for data mining, pretreat by data mining, including data formation, conversion, loss of data repair, so that converse raw data into valid data.

3. Principles of Intelligent Decision-Making of Public Data Warehouse

System sort the data of effective data warehouse by using data mining (such as: association rules, clustering, classification, decision trees, etc.) In order to adapted the truth of data increasing, data mining algorithms should also be correspondingly using data incremental data mining algorithms, effective knowledge produced by data mining are stored in the cloud center, provided to fixed and occasional users of system to use through the internet.

In this scenario, the data warehouse is the basis for intelligent decision-making process is the key. Therefore, to

construct effective method of data warehouse data validation to choose, but also to build more effective programs to achieve the incremental data mining, making mining results continue to improve the knowledge gained, so that increase knowledge of the universality.

4. Development Prospects of Intelligent Decision Support Technology of Public Data Warehouse Based on Cloud

The current data mining techniques are basically based on local enterprises, it is impossible to realize data mining globalization. With the development of cloud computing technology, this data mining technology will have a tremendous vitality.

Specific performance:

- (1) Data collection within the world, globalization of data mining to knowledge we got is difficult to achieve by any technical.
- (2) The development of cloud computing will lead to the development of this technology.
- (3) The user is a global technology service users and service a wide range.
- (4) This technology quickly and directly provide a wide range of services.
- (5) The technology areas covered by broad, such as medical diagnosis, natural disaster prediction and so on.
- (6) Development of network technology, making the portable user terminals to enhance and improve the range of users.

Information sharing is an inevitable requirement of mankind, the program can be across geographical and national boundaries, across language barriers, free access to the information needed to serve humanity.

5. Public Cloud-based Intelligent Decision Support Data Warehouse Technical Problems

The public cloud-based intelligent decision support data warehouse technology has broad prospects for development and development of the field, but this technology has some problems.

- (1) Privacy protection technology has led some users do not agree to provide the raw data. It may even provide a certain degree of privacy protection treatment.
- (2) Occasional users provide a lot of spam may increase the difficulty of the system noise-delete.
- (3) Because the data is stored in the cloud center, data security received a great threat.
- (4) The awareness of some companies may not support the technology to carry out.
- (5) How the costs of running the system to solve is a worthy subject of study.

Overall, the outlook for this program has a good momentum of development, but also the existence of defects can not be avoided, as technology development and a higher level of human knowledge, these problems will be properly resolved.

6. Conclusion

In this paper, as the development of cloud computing a public cloud-based data warehouse with intelligent decision support technology will inevitably arise, it will encourage the emergence of data mining technology, both in the applications or has a pole on the development prospects strong vitality, although this technology there are many constraints on their development, but it benefits a strong temptation, will become the future direction of development. Since the author is limited, I will be glad to discuss this issue with other researchers.

References

- Francesco Maria Aymerich, & Gianni Fenu. (2008). Simone Surcis. An approach to a cloud computing network. *First International Conference on the Applications of Digital Information and WebTechnologies*, 113-118.
- Gianni Fenu, & Simone Surcis. (2009). A cloud computing based real time financial system. Eighth International Conference on Networks. <http://dx.doi.org/10.1109/ICN.2009.71>
- Kuang Shenghui, & Li Bo. (2010). Cloud computing architecture and its application analysis. *Computer and Digital Engineering*, 38(3), 60-63.
- LuisM Vaquero, & Luis Rodero-Merino. (2008). Juan Caceres, *et al.* A break in the clouds: Towards a cloud definition. *ACM SIGCOMM Computer Communication Review*.
- Mladen A. Vouk. (2008). Cloud computing-issues, research and implementations. *Journal of Computing and Information Technology*, (4), 235-246.

Ni xianjun, Li guo, & Wu Yi Hui. (2007). based on cloud model of data mining technology. *Shandong Normal University (Natural Science Edition)*, 22(1), 4-48.

Wang Jiajuan, Lv Zhihui, Wu Jie, *et al.* (2010). Cloud computing technology development analysis and its application. *Computer Engineering and Design*, 31(20), 4404-4409.

Wang Lizhe, Tao Jie, Marcel Kunze, *et al.* (2008). Scientific cloud computing: Early definition and experience. *10 th IEEE International Conference on High Performance Computing and Communications*, 825-830.

Wang Peng. (2009). *Approached the cloud computing*. Beijing: People's Posts and Telecommunications Press.

Xiang Guibing, & He Jian. (2010). Cloud-based high performance distributed data mining methods. *Computer Engineering*, 36(5), 76-78.

Zhang Jianxun, Gu Zhimin, & Zheng Chao. (2010). Cloud computing research overview. *Computer Science*, 27(2), 429-433.

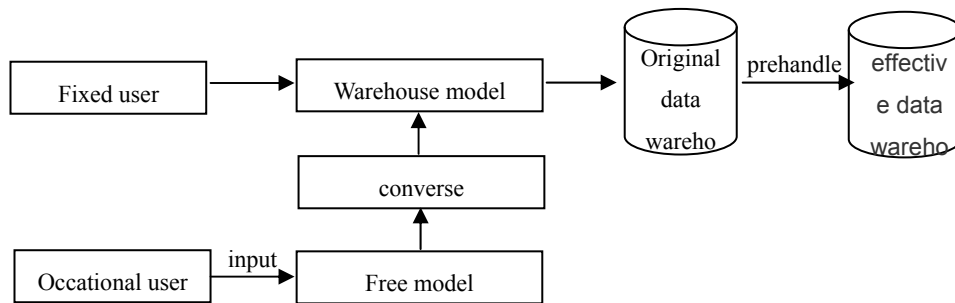


Figure 1. Constructure of original data warehouse

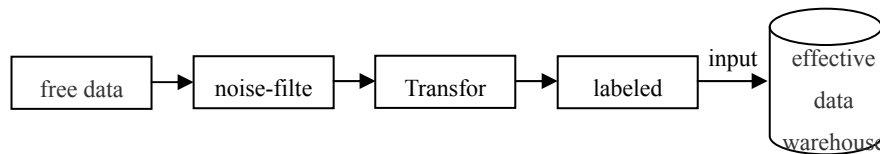


Figure 2. Constructure of effective data warehouse

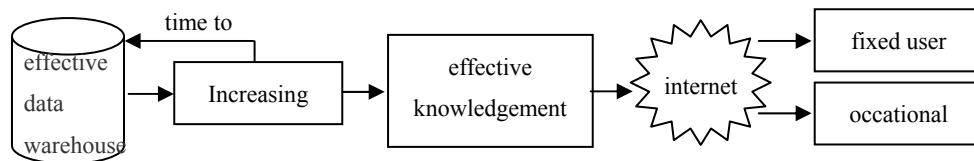


Figure 3. Data mining method

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.