Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 56 (2015) 111 – 118
Procedia Computer Science

The 12th International Conference on Mobile Systems and Pervasive Computing (MobiSPC 2015)

An Intelligent ETL Grid-Based Solution to Enable Spatial Data Warehouse Deployment in Cyber Physical System Context

Boubaker Boulekrouche[a], Nafaâ Jabeur[b,*], Zaia Alimazighi[a]

[a]*University of Sciences and Technologies, Houari Boumediene, BP 32 EL Alia, Bab Ezzouar Algiers, Algeria*
[b]*German University of Technology in Oman (GUtech), PO Box 1816, Athaibah PC 130, Muscat, Sultanate of Oman*

Abstract

Thanks to their spatially distributed sensors, Cyber-Physical System (CPS) applications are currently accumulating large amounts of heterogeneous data. When it comes to allowing several decision-makers to collaboratively plan their actions, these applications need appropriate tools for an efficient storage, analysis, and visualization of the available data. Spatial Data Warehouses (SDWs) have proven their efficiency in carrying out these operations. However, because of the increasing quantity of data, the Extract-Transform-Load (ETL) process (which is in charge of aggregating several data sources within a unified data storage repository) generally fails to update the SDW within predefined window times. In order to solve this problem, we propose in this paper to distribute the ETL tasks over a grid of computing resources. We also propose a multiagent-based approach that controls the ETL grid while allowing a convenient use of the shared resources. In addition to being the unique solution that uses grid computing for the ETL process of SDWs, our approach allows a joint use of archive and real-time data for personalized reporting and visualization of services envisioned to the decision-makers who are using the CPS application.

Peer-review under responsibility of the Conference Program Chairs

Keywords: Spatial Data Warhouse; Cyber-Phyiscal System, Multi-Agent System, Grids, Extract-Transform-Load (ETL)

1. Introduction

Cyber Physical Systems (CPSs) are being used in several application domains, including healthcare, environment monitoring, emergency systems, and intelligent transportation. They have recently emerged as promising tools where

*Corresponding author: Tel. +968-2206-1189

Email address: nafaa.jabeur@gutech.edu.om

Peer-review under responsibility of the Conference Program Chairs
doi:10.1016/j.procs.2015.07.176

the operations of the physical and engineered systems are monitored, controlled, coordinated, and integrated by means of a computing and communication core[1]. To this end, a variety of spatially distributed sensing devices are being used to collect *in-situ* data, anytime and anywhere, about application-related resources and events of interest. As a result, more efforts are needed to analyze and process the increasingly available data and reveal valuable information for the envisioned CPS applications. In several domains, these applications are expected to provide different decision-makers with a unified platform to collaboratively plan their actions about some ongoing events. This platform has particularly to deal with the challenging task of jointly using the heterogeneous large data sources of all decision-makers. Furthermore, for an enhanced collaboration, the platform should include mechanisms to deliver personalized services to decision-makers depending on their current requirements and the specificity of their operations.

The management of huge of amounts of data is being addressed in the research and development communities within two main fields; namely Big Data and Data Warehouses (DWs). On the one hand, Big Data provides solutions to store raw data which usage is not predetermined. On the other hand, DW technologies provide solutions to store cleaned, transformed, and semantically unified data for predetermined usage. Despite their high cost compared to Big Data solutions, DW solutions provide decision-makers with efficient tools for personalized reporting and visualization. In addition, since sensor data have mostly spatial components, DW, and more precisely Spatial Data Warehouse (SDW), is preferred to Big Data thanks to its long background in dealing with geo-referenced data. A SDW can be defined as a subject-oriented, integrated, time-variant, and non-volatile dataset including a collection of both spatial and non-spatial data in support of management's decision-making process[2]. A SDW is particularly characterized by its Extract-Transform-Load (ETL) process which is capable of merging data from different sources and creating specialized datamarts for a variety of applications. SDWs generally use SOLAP (Spatial On-Line Analytical Processing) tools for online processing along with Geographic Information System (GIS) functionalities for storing and visualizing spatial information[4].

With the increasing availability of data within the context of CPS applications, the ETL process, which is already time-consuming, faces problems in carrying out the necessary processing operations within acceptable timeframes[5]. Furthermore, although some approaches[7,8] have been proposed to improve the freshness of data in the DW, current ETL processes are still unable to adequately combine the use of real-time data along with the archive data (stored in data sources). The common idea of reducing the processing load by purchasing additional computing resources is being seen as an expensive option[5]. Alternatively, moving to a parallel, distributed model that reuses the resources available with each of the decision-makers could be well suited to the context of SDW[9]. This model is found in the grid computing paradigm[10].

Grid computing has attracted extensive research and development works, particularly for its easiness and inexpensiveness of adding new processing nodes to the infrastructure[11]. In addition to offering mechanisms to enable the reallocation of critical DW resources for use by data mining and reporting tools, grid computing can also drive significant benefits by improving information access and responsiveness as well as adding flexibility, which are all important components of solving the DW dilemma[9,12]. While several studies have proposed to distribute DWs in grid environments[13,14], only one research work[5] has proposed to balance the tasks of the ETL process over a grid. In addition to contrasting the existing works by addressing the context of SDW, this paper proposes a multiagent -based approach to distribute Spatial ETL (SETL) processes over a grid. Our approach is also capable of enabling a joint use of real-time and archive data within the context of SDWs for CPS applications.

In the remainder of the paper, Section 2 gives an overview of existing works that have used grid-based approaches for SDWs. Section 3 presents our multiagent-based solution to improve the SETL process by distributing its tasks over a grid of computing resources. Section 4 presents the current state of the implementation of our approach.

2. Related works

SDWs result from the integration of GIS technologies and data warehousing capabilities. They are being used in a wide range of geospatial application domains, including business, military, industry, and disaster management. A typical SDW architecture (Figure 1) includes an Extract-Transform-Load (ETL) that collects data from multiple data sources (extraction), cleanses and normalizes this data to meet some integrity standards of the warehouse (transformation), and then brings the resulting data into the warehouse as new records (load). Once the data is loaded, analysis can be performed by querying the SDW directly or, most often, smaller data subsets (datamarts) are created

for specific uses and then queried separately. Third-party tools (e.g., SOLAP tools) are commonly used to analyze the data in the SDW and the datamarts and generate customized reports for visualization purposes.

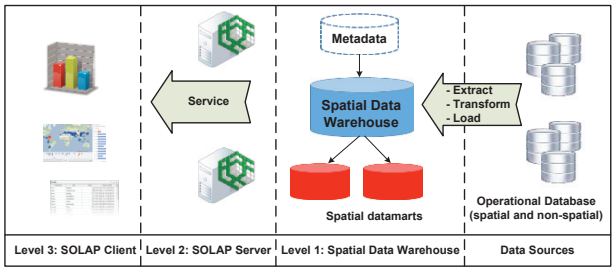

Fig. 1. SDW architecture.

Several research works have designed SDWs for specific context of use, such as managing petroleum resources[23], managing epidemics[23], and discovering ecological patterns and processes[25]. Other research works have proposed general conceptual models[18,19], logical models[20,21], and implementation strategies and optimization techniques[2,22]. The SDWs proposed in these works are basically designed to support the needs and the perceptions of one single actor, although probably at different scales. In critical scenarios (e.g., hazard management), several decision-makers (e.g., police authority, medical core, municipality services, etc.) generally need to make collaborative decisions to optimize their actions and use of resources. Commonly, this collaboration process is based on the use of multiple individual SDWs or on the use of one SDW supporting a single perception only. Both approaches do not always lead to easy agreements on concepts, events' assessments, and action priorities. Therefore, there is a high need for a common infrastructure based on a federated SDW[9] that integrates the different perceptions of decision-makers and allows customized analysis and visualization of data.

In order to implement the expected federated SDW, several processing steps must be performed. For instance, data coming from several sources must be analyzed, filtered, cleansed, and transformed to meet specific requirements in terms of scale, format, detail levels, and quality. This time-consuming processing is facing an increasing complexity because of the growing availability of data which needs to be processed within predefined window times. Instead of acquiring new powerful hardware to decrease the processing time, the less expensive solution consisting in reusing the available, distributed computing resources within the context of grid computing has been identified as a promising alternative[5,9,12,13]. Indeed, grid computing provides an efficient approach to harnessing distributed resources, while promoting scalability, reliability, cost saving, and better throughput[12].

Several studies have proposed to distribute DWs in a grid environment[13,15,16], with a particular emphasis on query distributed approaches[14,17]. In this regard, Akinde et al.[26] have discussed query processing in an environment consisting in distributed DWs. In this environment, one site plays the role of coordinator, whereas most of the processing loads being performed at local sites. Costa and Furtado[13] have investigated a Grid Data Warehouse Parallel Architecture (Grid-Dwpa) as an efficient architecture to deploy large DWs in grids with high availability and good load balancing. In addition to a data allocation and a partial replication strategies, the authors have proposed scheduling solutions that maximize performance and throughput of the grid-enabled architecture for OLAP. Costa and Furtado[27] have presented a scheduling approach for efficient query processing in the Grid-Dwpa environment. The system generates site and node tasks, forecasts the necessary time to execute the task at each local site, estimates total execution times, and assigns task execution to sites accordingly. Another work on grid-aware DWs[28] has considered the scenario where the data of a single organization is distributed across a number of operational databases at remote locations. Each operational database has capabilities for answering OLAP queries and accessing to a possible variety of other computational and storage resources which are located close by. In all these works, the grid computing concept has been used to implement distributed DW only. They did not support spatial data which are particularly characterized by their complex coordinates, different formats, and topological relations. Furthermore, to the best of our knowledge, the important potential of grid computing was not appropriately used to alleviate the heavy-duty ETL process. We, therefore, believe that additional research and development efforts are needed to integrate the grid computing

techniques within the domain of SDWs while focusing on ways to lighten the complex ETL process in order to allow convenient processing within tolerated window times.

3. Toward the improvement of the Extract-Transform-Load process

Starting from the firm conviction that an adequate use of grid computing could appropriately alleviate the ETL process in handling the increasing amounts of heterogeneous data, we examine in what follows how to distribute the tasks of the ETL over distributed resources. We also study the use of the multiagent system paradigm in enabling an intelligent control and coordination of the ETL grid.

3.1. Grid-Based SETL

The ETL process includes several operations to be applied to the different data sources. The complexity of these operations are particularly affected by the format and quality of the data sources as well as the degree of their heterogeneity. The distribution of the ETL processes over a grid could be done according to several approaches. A possible approach consists in clustering the pool of shared resources into three groups dedicated to the operations of extraction, transformation, and loading respectively. This solution increases the communication between the computing resources in the grid. However, it is fault-tolerant thanks to the redundancy of resources. In this paper, we adopt a slightly different approach where every computing resource is able to carry out the whole ETL process, without being obliged to complete this process every time. For instance, depending on its current situation (e.g., availability and processing capabilities), a given computing resource could be assigned the task to execute the operations of extraction and transformation and then the loading task is assigned to another computing resource. In addition to increasing fault-tolerance mechanisms compared to the first solution, this approach gives more comfort to the decision-makers who will be able to apply the whole ETL process on their own computing machines without relying on the computing resources of partners. Based on this approach, we propose an architecture (Figure 2) where we extend the ETL process to SETL (Spatial ETL). In the proposed configuration, the clients of the grid are the different data sources which are concurrently requesting to load their data into the SDW.

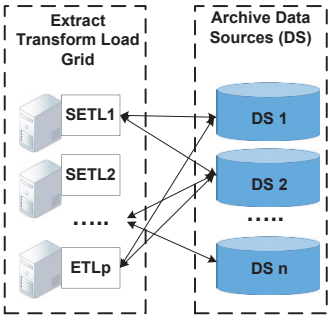

Fig. 2. SETL (Spatial Extract-Transform-Load) grid

Basically, the different decision-makers need to load updates from their data sources into the SDW at sporadic moments depending on their activities. To this end, the data sources may need to compete for the shared resources. This competition ultimately aims to adequately and fairly distribute the SETL tasks over the grid. The issue of balancing tasks over the grid has been studied intensively in the literature, especially over parallel architectures, with a variety of approaches, including priority-based[30], Fuzzy-logic[31], profile-based[32], and agent-based[33]. Furthermore, some attempts have addressed the grid load balancing issue in the context of DW[13], but without supporting spatial data. In the next section, we address the issue of monitoring the SETL grid within the context of SDWs using an intelligent approach.

3.2. Intelligent Monitoring of the SETL Grid

As already stated, there is lack of research in the area of applying grid computing to the SETL process. In order to contribute to this domain, a very structured workflow of operations must be defined. This workflow will particularly determine the right time to start the SETL process, which shared resources will carry out the required processing tasks and in which sequence, and when the transformed data will be loaded in the SDW. Because of the complexity of processing as well as the varying requirements of decision-makers due to some unpredictable situations, an adequate solution supporting distributed operations within a competitive environment is needed. The multiagent system paradigm was adopted in this paper because of its proven flexibility, autonomy, and intelligence to solve complex problems within highly dynamic, constrained, and uncertain environments[3]. To this end, we propose an architecture (Figure 3) where an agent-based intelligent module is used. This module has several tasks that we outline in what follows.

Sending data to the archive. Since huge amounts of data are being gathered by the spatially distributed sensors of the CPS, there is a need to process this data and load it in the SDW at convenient times. Because of the heavy processing loads of the SETLs as well as the need to respond in timely fashion to the decision-makers, it is not necessary to overload the system with processing all data immediately. The intelligent module is then used to review the current needs of decision-makers and decide whether the new data has to be sent straightaway to the Real-Time Integration Module (RTIM) or has to be scheduled for SETL processing. For immediate considerations, the RTIM receives a copy of the data and carries out some processing to create specific contents as per the requirements of decision-makers (details of these processing are out of the scope of this paper). The original data will then be scheduled by the intelligent module for subsequent SETL processing.

Balancing the load in the SETL grid. The requirements of the different decision-makers are generally highly affected by ongoing events. Concurrent requirements to the distributed SETLs could reach pick-moments in critical situations, where decision-makers need to push their archive data (data sources) and real-time data (from sensing devices) to the SDW. The agent-based module has then the responsibility to assign tasks to the SETLs based on the priority of requests, ongoing events, and any prior agreement made between the decision-makers sharing the same CPS application (e.g., priority is given to the owner of the computing resource). To this end, the module receives frequent updates on the current processing and available computing resources from the SETL grid. Actually, by sending data to RTIM, the intelligent-module undertakes the task of balancing the grid load since some processing could be delayed. The intelligent module also prepares a backup plan for data processing in case of failure of any computing resource.

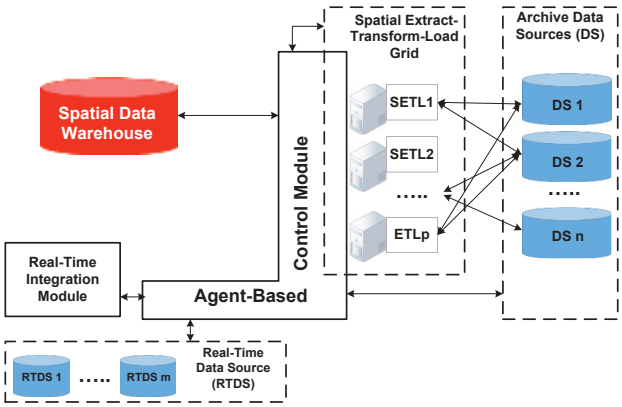

Fig. 3. Integration intelligence in the SETL process

Coordinating the data loading process. As the SDW is a shared data repository, the agent-based control module has the duty to coordinate the concurrent update of the SDW by the distributed SETLs. A priority-based mechanism is used to ensure an efficient and non-conflicting update of the SDW.

Facilitating the collaboration of decision-makers. The SETL grid as well as the SDW are shared resources between the different decision-makers. With appropriate management strategies, the agent-based control module could increase

the availability of these resources and avoid redundant operations, allowing thereby the different decision-makers to jointly plan their actions without phase shifts. To this end, a process of advertisement and discovery can be used along with performance prediction mechanisms, such as Performance Analysis and Characterize Environment (PACE)[32].

3.3. Multiagent System Architecture of the SETL Grid Control Module

In this section, we outline the architecture of our multiagent-based system for the control and monitoring of the SETL process (Figure 4). Our system includes a *Broker Agent* that receives frequent notifications for the availability of new real-time data. Depending on the current situation and the services being requested by the decision-makers, the *Broker Agent* may decide, as stated earlier, to send the data to the RTIM for their immediate consideration in the current decision-making process. The *Agent Broker* also receives requests from the data sources to schedule the loading of some of their pending data into the SDW. In this case, the request is sent to a *Scheduler Agent* that carries out the necessary operations to schedule the processing over the SETL grid in a balanced way. To achieve this task, *the Scheduler Agent* may consult with a *Prioritizer Agent* that frequently updates the priority of tasks as per the requirements of the decision-makers and the current situations. The *Scheduler Agent* also gets information on the current available resources as well as the current processing activities in the grid from the *Tracker Agent*. This latter agent frequently receives updates from the *Local Manager Agents* which are located in the SETL machines. Once a balanced schedule for SETL processing activities is ready, the *Scheduler Agent* instructs a *Backuper Agent* to identify backup resources and prepare an alternative plan in case of failure of any SETL machine in the grid. Details about the different agents as well as their communications and processing will be the subject of a future publication.

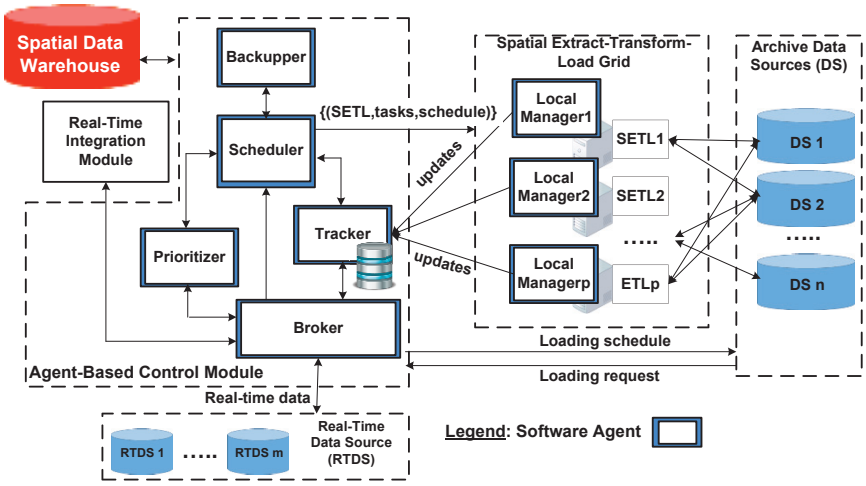

Fig. 4. Multiagent system architecture

4. Ongoing Implementation

For the sake of illustration, we have considered the scenario of monitoring industrial accidents in Algeria. In this scenario, several decision-makers, including industrial risk experts, civil protection managers, and public safety officers, are collaborating to monitor an ongoing accidents (e.g., gas leak). The SDW, which is basically integrating the data sources of all stakeholders, is expected to be queried and analyzed with appropriate tools in order to accommodate experts with geo-decisional information according to their specific needs as well as their perceptions of the current industrial accidents. To this end, we developed a decision-support system called GéOLAP (Figure 5) which is implemented using ROLAP server Mondrian and MapServer tools for the processing and analysis of non-spatial data and spatial data respectively. The SDW is implemented with the open-source tools PostGIS and PostgreSQL. At the client side, the visualization of non-spatial data is performed using OLAP client Jpivot whereas the visualization of spatial data is carried out with the JavaScript OpenLayers tool. The management of interaction and communication

tasks between OLAP Mondrian serveur and MapServer is monitored by a Processing and Personalization Module. This module is particularly responsible for reporting SDW data according to the specific needs of decision-makers as well as integrating the real-time data received from the RTIM (see Section 3.3).

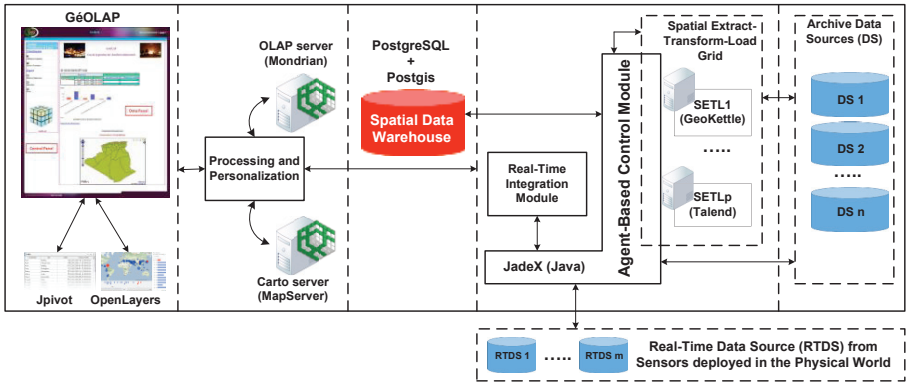

Fig. 5. GéOLAP system configuration

The SDW population is done from one single SETL machine using the FME software tool. We are currently investigating the implementation of the SETL grid with several technologies and platforms, including GeoKettle and Talend. The software selected for the implementation of the multiagent system is the Java-based JadeX platform. The Processing and Personalization Module and the Real-Time Integration Module are both being developed in Java. Performance and additional details about the implementation and the performance of our system will be the subject of an upcoming research paper.

5. Conclusion and future works

In this paper, we addressed the growing important issue of managing huge amounts of data within the context of the emergent filed of Cyber Physical System (CPS). We proposed a new approach that enables the integration of heterogeneous spatial and non-spatial data within a unified Spatial Data Warehouse (SDW). The SDW is ultimately used to facilitate the collaboration between several decision-makers on some ongoing events of common interest. Unlike any existing solution, we proposed to distribute the Spatial Extract-Transform-Load (SETL) tasks over a grid of computing resources. We also proposed to implement a multiagent-based solution to adequately schedule the processing activities over the grid. Our multiagent-based solution is also intended to allow a joint use of real-time and archive data within the same CPS application.

Currently, our prototype is able to report adequately personalized spatial and non-spatial data to different decision-makers within the context of monitoring industrial accidents in Algeria. In order to improve its performance, ongoing efforts are aiming to adequately set up the pool of shared computing resources and implement the intelligent solution for its management.

References

1. Rajkumar, R., Lee, I., Sha, L., Stankovic, J. (2010). Cyber-physical systems: the next computing revolution. Proceedings of the 47th Design Automation Conference, ACM, 2010
2. Stefanovic N, Han, J., Koperski, K. (2000). Object-Based Selective Materialization for Efficient implementation of Spatial Data Cubes. IEEE Transactions on Knowledge and Data Engineering, Vol. 12, n° 6, pp. 938-958
3. Bandyopadhyay S., Coyle E.J. (2013). An energy efficient hierarchical clustering algorithm for wireless sensor networks", Proc. of INFOCOM 20013, IEEE Societies, 2013, vol. 3, pp. 1713-1723
4. Raffaetà, A., Leonardi, L., Marketos, G., Andrienko, G., Andrienko, N., Frentzos, E. (2011) Visual mobility analysis using T-warehouse. International Journal of Data Warehousing and Mining, 7(1), pp. 1–23. Doi=10.4018/jdwm.2011010101

5. Santos, V., Oliveira, B., Silva, R., Belo, O. (2012). Configuring and executing etl tasks on grid environments - requirements and specificities, Procedia Technology, Volume 1, 2012, Pages 112-117, ISSN 2212-0173, http://dx.doi.org/10.1016/j.protcy.2012.02.022
6. Yagoubi, B., Meddeber, M. (2010). Distributed Load Balancing Model for Grid Computing. ARIMA journal, vol. 12, 2010, pp. 43-60
7. JinGang, Sh., Bao, Y., Leng, F., Yu, G. (2009). Priority-Based Balance Scheduling in Real-Time Data Warehouse. HIS, 2009, Hybrid Intelligent Systems, International Conference on, Hybrid Intelligent Systems, International Conference on 2009, pp. 301-306
8. Thomsen, C.; Pedersen, T.B.; Lehner, W. (2008). RiTE: Providing On-Demand Data for Right-Time Data Warehousing. ICDE 2008. IEEE 24th International Conference on Data Engineering, vol., no., pp.456,465, 7-12 April 2008
9. Butte, B. (2004). Solving the data Warehouse dilemma With grid technology", IBM Global Services, Aug. 2004, IBM. Available at: http://csis.bits-pilani.ac.in/faculty/goel/course_material/Data%20Warehousing/I%20sem%202005-06/Assignemt%202/GW510-5041-00F.pdf
10. Foster, I., Kesselman, C., Tuecke, S. (2001). The Anatomy of the Grid: Enabling Scalable Virtual Organizations. International Journal of High Performance Computing Applications 15, pp. 200-222
11. Demiya, T., Yoshihisa, T., Kanazawa, M. (2008). Compact grid: a grid computing system using low resource compact computers. Int. J. Commun. Netw. Distrib. Syst. 1, pp. 17
12. Zode, M. (2008). "Grids in Data Warehouses", available at: http://www.tdan.com/view-articles/9378 Last access April 25, 2015
13. Costa, R. L. d. C., Furtado, P. (2007). An SLA-Enabled Grid DataWarehouse. In: 11th International Database Engineering and Applications Symposium (IDEAS 2007), pp. 285-289. Banff, Alberta, Canada
14. Wehrle, P., Miquel, M., Tchounikine, A. (2007): A Grid Services-Oriented Architecture for Efficient Operation of Distributed Data Warehouses on Globus. In: Advanced Information Networking and Applications, 2007, pp. 994-999
15. Iqbal, S., Bunn, J. J., Newman, H. B. (2003). Distributed Heterogeneous Relational Data Warehouse In A Grid Environment. In: Computing in High Energy and Nuclear Physics, La Jolla, California
16. Dubitzky, W., McCourt, D., Galushka, M., Romberg, M., Schuller, B. (2004). Grid-enabled data warehousing for molecular engineering. Parallel Computing 30, pp. 1019-1035
17. Pascal, W.: A Model for Distributing and Querying a Data Warehouse on a Computing Grid, pp. 203-209.(2005)
18. Ezzeddine S., Turki S.Y., Faiz S. (2014). Enriching Dimension Hierarchies with Topological Relations to improve the development of Spatial Data Warehouse. The 6th International Conference on Advances in databases, Knowledge, and Data Applications (DBKDA'2014), Chamonix, France, Avril 2014, pp. 35-40
19. Fidalgo, R., Cuzzocrea, A. (2012) An Enhanced Spatial Data Warehouse Metamodel", in CAiSE Forum, Vol. 855 of CEUR Workshop Proceedings, pp.32-39, 2012
20. Gascueña, C., Guadalupe, R. (2009). A multidimensional methodology with support for spatio-temporal multigranularity in the conceptual and logical phases. In Taniar, D. (Ed.), Progressive methods in data warehousing and business intelligence: Concepts and competitive analytics (pp. 194–230). Hershey, PA: IGI Global
21. Zaamoune, M., Bimonte, S., Pinet, F., Beaune, P. (2013). A new relational spatial OLAP approach for multi-resolution and spatio-multidimensional analysis of incomplete field data. ICEIS 2013 IN-STICC International Conference on Enterprise Information Systems, Jul 2013, Angers, France.p. 145 - p. 152
22. Rao, F., Zhang, L., Yu,X., Li,Y., Chen, Y. (2003). Spatial hierarchy and OLAP-favored search in spatial data warehouse. Proceedings of the 6th ACM International Workshop on Data Warehousing and OLAP, DOLAP '03 (pp. 48-55)
23. Li, Y., Wang, L., Ji, L., Liao, Ch. (2013). A Data Warehouse Architecture supporting Energy Management of Intelligent Electricity System. In Proc. of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013), doi:10.2991/iccsee.2013.177
24. Cembalo, A., Ferrucci, M., Pisano, F.M., Pigliasco, G. (2013). SOLAP4epidemiologist: A Spatial Data Warehousing Application in Epidemiology Domain. In the Proc. of the 15th International Conference, DaWaK 2013, Prague, Czech Republic, August 26-29, 2013, pp. 97-109
25. McGuire, M., Gangopadhyay, A., Komlodi, A., Swan, Ch. (2008). A user-centered design for a spatial data warehouse for data exploration in environmental research, Ecological Informatics, Volume 3, Issues 4-5, October 2008, Pages 273-285
26. Akinde, M. O., Bhlen, M. H., Johnson, T., Lakshmanan, L. V. S. and Srivastava, D. (2003). Efficient OLAP query processing in distributed data warehouses, Information Systems 28, pp. 111-135, Elsevier
27. Costa R. and Furtado P (2008). Optimizer and QoS for the Community Data Warehouse Architecture, in New Trends in Database Systems: Methods, Tools, Applications", Eds. D. Zakrzewska, E. Menasalvas, L. Byczkowska-Lipiñska1, Springer-Verlag, 2008.
28. Lawrence M. and Rau-Chaplin A. (2006). The OLAP-Enabled Grid: Model and Query Processing Algorithms" in Proceedings of the 20th International Symposium on High Performance Computing Systems and Applications (HPCS'06), IEEE, Eds. R. Deupree, St. Johns, Canada, May 2006
29. A. R. Mury, B. Schulze, and A. T. A. Gomes. (2010). Task distribution models in grids: towards a profile-based approach. Concurrency and Computation: Practice and Experience, 22(3), 358-374. doi: 10.1002/cpe.1474
30. Kumar, S., Singhal, N. (2012). A Priority based Dynamic Load Balancing Approach in a Grid based Distributed Computing Network. International Journal of Computer Applications (0975-8887), Vo. 49, No.5, July 2012, pp. 11-13
31. Helmy, T., Al-Jamimi, H., Ahmed, B., Loqman, H. (2012). Fuzzy Logic–Based Scheme for Load Balancing in Grid Services. A Journal of Software Engineering and Applications, 2012, 5, 149-156 doi:10.4236/jsea.2012.512b029
32. Nudd, G., Kerbyson, D., Papaefstathiou, E., Perry, S., Harper, J., Wilcox, D. (2000). Pace - A Toolset for the Performance Prediction of Parallel and Distributed Systems. Int. J. High Perform. Comput. Appl., 14(3): 228–251, 2000, ISSN 1094-3420