Class Project Book Chapter
Innovations in Information System
Series: 3 | Year: 2017 | ISBN: 978-967-0194-97-4

11
USING SOCIAL NETWORK'S DATA BY EXTRATION TRANSFORMATION LOADING (ETL)

Fatemeh Laghaei, Othman Bin Ibrahim

11.1 INTRODUCTION

During the present time one of the objectives of a successful business strategy is to obtain and keep the most profitable customers [3]. To enhance our business such as marketing or Customer Relationship Management (CRM), it is important to analyze customers' communication patterns based on customers' social ties.

In the era of social customer, the customer seizes control of the business ecosystem and the business value of community and collaboration in the new world of the social customer cannot be underestimated [2].

Nowadays, social networks become part of our life and using them for different purposes such as communication and business to find value. Social networks vendors state a number of business advantages, such as using online communities as a source of high-quality leads and a vehicle for crowd sourcing solutions to client-support problems [6]. Due to a power of social network websites in business and enterprise, it causes a dramatic value and made integration of existing Enterprise Information System (EIS) with social network's data. However, it is a challenging task to use social network's data which needs a new requirement from

enterprises. This chapter investigates on data and its format of social networks besides type and structure of EIS in different level of integration methods in order to purpose a proper model for integrating them.

11.2 DATA INTEGRATION LEVELS

Data integration, application integration, business process integration, and user interaction integration are the four different levels of enterprise business integration that take place in an information technology system [10].

Data Integration: This level proposes an integrated view point of the enterprise data that is sprinkled all the way through enterprise. This combined outlook can be established by using a mixture of unlike methods and technologies. In physical view the data has been enclosed from different data sources and merged into an integrated data warehouse or operational data store. An example can be like merging client data from a CRM database into an ERP database [10].

Application Integration: Application integration offers a combined view point of business applications that reside within or outside an enterprise. Observing, managing and organizing the flow of events like transactions, messages, or data between different enterprise applications resulted in combined outlook. There is a variety of technologies to implement application integration relying on the requirements of an integration project [10].

Business Process Management: Business process management gives a united view of an enterprise processes. Developers and programmers can evaluate, model, and simulate business processes and their basic activities more easily by using business integration software. The benefit of business process management is that the

design features of business process analysis and design are protected from physical business process management and application accomplishment thoughts [10].

User Interaction Integration: User interaction integration provides users with a single modified and secure interface to the business content like processes, applications, and data which these are required for doing their jobs. The interface also allows users to work together and share data with each other. An enterprise portal is an example of a product that supports user interaction integration [10].

11.3 DATA INTEGRATION TECHNOLOGIES

There are several technologies of data integration that address the integration based on use of data, volume of data, and the number of data users. Following sections are intended to describe these technologies. At the end of this section, Table 1 has provided a comparison study of them.

11.3.1 Transforming Extracting and Loading

ETL stands for Extraction, Transformation and Loading. ETL is a process that involves the following tasks: (see Figure 11.1)

(a) **Extracting** data from multiple heterogeneous systems or different sources or archive systems
(b) **Transforming** the data - which may involve cleaning, filtering, validating and applying business rules
(c) **Loading** the data into a data warehouse or any other database or application that houses data

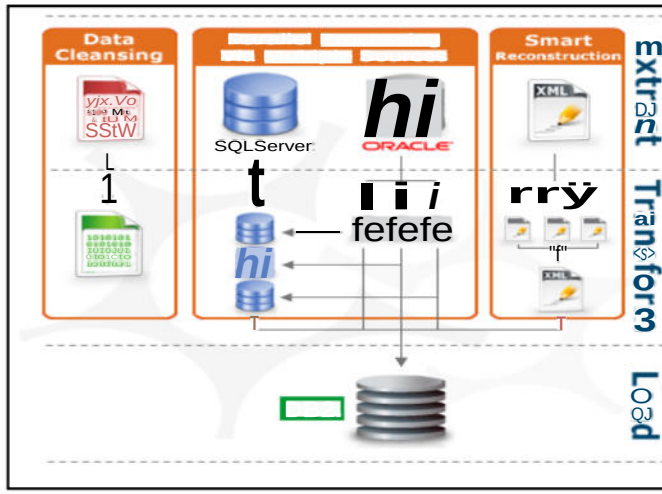

Figure 11.1 ETL diagram

In [12] an ETL process is the foundation of data warehouse. A well-designed ETL process extracts data from the data sources, enforces data quality standards, so that data can be used by applications developers and end users can make strategic decisions. That is, the data are extracted from source systems, which undergo a sequence of transformations before they are loaded into the data warehouse. The repository systems that contain data sources for data warehouse can range from spreadsheets to systems mainframe. The transformations process are usually complex, implemented in procedural programs, either outside databases such as (in C, Java, Pascal, etc.) or within the database. Design of an ETL process is usually composed of six tasks defined in [8]: (i) select data for extraction (ii) define the data source (usually from various heterogeneous sources), (iii) transforming the sources, (iv) joining sources, (v) select the destination for load (v) join attributes of the sources of data attributes destination, (vi) load the data into the data warehouse.

The ETL system adds significant value to data. It is far more than plumbing for getting data out of source systems and into

the data warehouse. Specifically, the ETL system:
(a) Removes mistakes and corrects missing data
(b) Provides documented measures of confidence in data
(c) Captures the flow of transactional data for safekeeping
(d) Adjusts data from multiple sources to be used together
(e) Structures data to be usable by end-user tools

ETL is both a simple and a complicated subject. Almost everyone understands the basic mission of the ETL system: to get data out of the source and load it into the data warehouse. And most observers are increasingly appreciating the need to clean and transform data along the way; So much for the simple view. It is a fact of life that the next step in the design of the ETL system is to break into a thousand little sub cases, depending on your own weird data sources, business rules, existing software, and unusual destination-reporting applications [12].

11.4 ENTERPRISE APPLICATION INTEGRATION

EAI stands for Enterprise Application Integration. Enterprise Application Integration is the global process of integrating a company's data across its various application systems and thereby enabling these systems to talk to one another [11]. Typically EAI is a code based package providing application-to-application connectivity with standard interfaces. EAI does not perform any data oriented integration [14]. Table 11.1 illustrates the comparison between these two technologies.

Table 11.1 Comparison of EAI and ETL [11]

	EAI	**ETL**
Focus	Operational Apps	Data Warehousing
Timing	Real-Time	Batch
Data	Transactional	Historical
Metadata	Limited • Message Metadata	Rich • Dimensional Metadata
Transformations	Format Oriented • Code Supported	Analytic • Joins • Aggregations • Concatenations
Volumes	Throughput • Single Transactions • Messages/Seconds(Kb)	Size • Days or Weeks of Data • Records Per Min(Gb)

As above Table 1 shows there are some similarities and differences between these two technologies which are:

(a) **Similarities-** data integration from disparate systems (mapping form disparate sources), both technologies are based on the concept of a 'unified view'.

(b) **Differences**- the purpose, speed, direction and amount of data that are transformed and 'placed' within the unified view from the external sources, EAI is the connector between systems, whereas ETL is a data warehousing process performed in an environment that is separate from the systems themselves. EAI systems retrieve small amounts of data while ETL systems retrieve large amounts of data (batches) of data from different sources [11].

Based on this comparison, EAI cannot handle social networks 'data integration which is in other environment and need batch loading. Besides, the integration in application level for variety of external systems which change rapidly and continuously

is impossible or high costly.

11.5 MASTER DATA MANAGEMENT AND CUSTOMER DATA INTEGRATION

Customer Data Integration (CDI) and Master Data Management (MDM) are getting significant buzz in both information technology (IT) and business circles. MDM represent the data management disciplines and processes, and CDI represents the authoritative systems for customer data. CDI is a subset of MDM, which can be considered the overall discipline of managing information domains across an enterprise. Thus, CDI is specific to automating the management and reconciliation of customer data, while MDM involves maintaining the master data across the domains, including product, chart of accounts or inventory.

Following Table 11.2 describes the differences between data integration technologies which have explained above, based on data movement, processing, interface type, and integration method [14].

Table 11.2 Comparison of data integration technologies [14]

Integration Technology	Data movement	Processing	Interface Type	Integration Method
EAI	None: In application messages	Online – event oriented	Static: written as an application	Does not support data integration
ETL	Table rows or flat file records	Batch	Static: SQL or I/O	Static sources with file or DBMS access
MDM	PARTY/CUSTOM ER Unique ID and details	Online – transaction oriented	SOA	Highly flexible based on subject areas

Based on this comparison and our integration requirement which is high amount of data from various social networks in data

layer, ETL is the best technology for integrating social network's data and EIS in organizations.

11.6 PROPOSED DATA INTEGRATION TECHNOLOGY

During the present time the interactions between an individual user and the communities have presented by social networks. On the other hand, The ETL system adds significant value to data. The ETL system removes mistakes and corrects missing data, providing documented measures of confidence in data, capturing the flow of transactional data for safekeeping, adjusting data from multiple sources to be used together, structuring data to be usable by end-user tools and also the complicated interactions between different systems is provided by ETL as the same with the intelligence in the areas of content and relationships [12]. ETL processes bring together and combine data from multiple source systems into a data warehouse or EIS, enabling all users to work off a single and integrated set of data. As a result, an organization has no longer to collect data or argue about which data is correct, but one that uses information as a key process enabler and competitive weapon [1].

Thus, due to the significant growth of both users and communities, ETL is used as a bridge between the data of user and also as a link between business databases. Moreover, capturing data and intelligence provided business value regarding to search and advertising analytics, thus real time business intelligence needs to be updated continuously to support the ongoing electronic conversation. In addition, one registration access to many communities in social networks represent the considerable business value to them because of the exponential potential of more users using more content with more relationships. ETL can help by accepting any format, any log-in, for greater reach. Another issue to propose ETL is that the social networks do not pose any restriction in terms of the amount of data which a user can publish over the web. Since massive amount of data is

available on the network, integration of them is a crucial issue. So the need for ETL arises from extracting data from all of the disparate source systems and running large volumes of this data through the mappings that are developed with minimal impact on the operating systems.

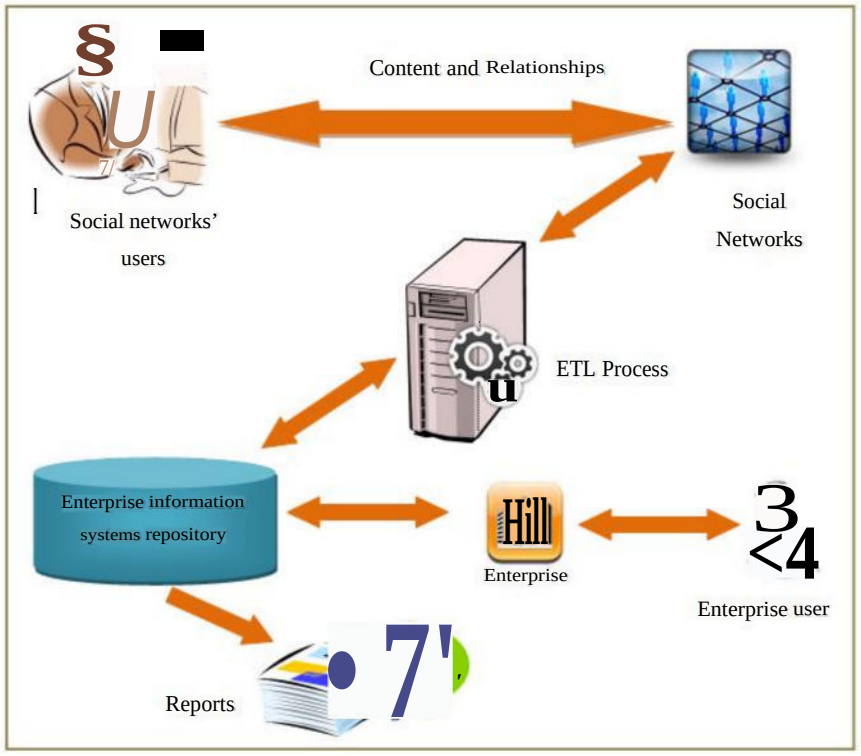

Figure 11.2 Integrating social network's data by ETL processes

Figure 11.2 shows as users interact with communities, intelligence is created. That is captured as intelligence with ETL as the engine, which is shared under many different business rules either in content management or community value. This is then either mined for reports to all aspects of the social network enterprise or shared with other business with their own set of data and users.

11.7 CHAPTER SUMMARY

The goal of this chapter is to provide the reader with an overview of findings regarding the convincingness of integration of social network's data and enterprise information systems and to reflect possible explanatory data integration technologies. To achieve this goal, the authors have summarized the comparison among data integration technologies. This chapter has examined and looked into several related concerns and research fields to provide a more explicable and important approach for using the advance of social network in enterprise, and trying to find the best integration technology to integrate social network's data and information systems for organizations to obtain more benefits.

REFERENCES

[1] Wayne Eckerson and Colin White, *Evaluating ETL and Data Integration Platforms*, The Data Warehousing Institute (TDWI), 2003.

[2] Paul Greenberg. *CRM at the Speed of Light: Social CRM Strategies, Tools, and Techniques for Engaging Your Customers.* Fourth Edition. McGraw-Hill. 2009.

[3] H. Hwang, T. Jung, and E. Sub, *An ltv model and customer segmentation based on customer value: a case study on the wireless telecommunication industry*. Expert Systems with Applications. 2004.

[4] S. Chawathe, H. Garcia-Molina, J. Hammer, K. Ireland, Y. Papakonstantinou, J. Ullman,and J Widom. *The TSIMMIS project: Integration of heterogeneous information sources*. In Proceedings of the 10th Meeting of the Information Processing Society of Japan, pages 7–18, Tokyo, Japan, October 1994.

[5] Boanerges Aleman-Meza, Meenakshi Nagarajan, Cartic Ramakrishnan, Li Ding, Pranam Kolari, Amit P. Sheth, I. Budak Arpinar, Anupam Joshi, Tim Finin. *Semantic*

Analytics on Social Networks: Experiences in Addressing the Problem of Conflict of Interest Detection. May 23–26, 2006. the International World Wide Web Conference Committee (IW3C2). ACM 1-59593-323-9/06/0005.

[6] DestinationCRM.com (2009). *Who Owns the Social Customer?, CRM* magazine's in-depth report on the state of social media in CRM, June 2009, AT&T Labs-Research Technical Report.

[7] Yutaka Matsuo, Junichiro Mori, and Masahiro Hamasaki .*POLYPHONET: An Advanced Social Network Extraction System from the Web.*

[8] J. Trujillo, S. Luján. *A UML Based Approach for Modeling ETL Processes in Data Warehouses*, in 22nd International Conference on Conceptual Modeling. 2003: USA, Chicago, 307-320.

[9] M.R. Genesereth, A.M. Keller, and O.M. Duschka, Infomaster: *An information integration system.* In Proceedings of 1997 ACM SIGMOD International Conference on Management of Data, pages 539–542, Tucson, Arizona, May 1997.

[10] Denish Gupt. *The levels for successful enterprise business data integration process*, FEB 2011.

[11] Source: http://www.crm2day.com/content/t6_librarynews_1.php?id=50036

[12] R. Kimball, J. Caserta. *The Data Warehouse ETL Toolkit*. 2004: Wiley Publishing.

[13] Jarvenpaa, S.T., L.N. M., *Vitale, Consumer trust in an Internet vendor.* Information Technology and Management 2000. 1(1&2): p. 45-71.

[14] Franchise Tax Board's (FTB) Enterprise Architecture (EA) Data Center of Excellence. *Data Integration Strategy – v5.0.* Des 2008.