

A MODEL FOR INTEGRATING SOCIAL NETWORK'S DATA AND  
ENTERPRISE INFORMATION SYSTEMS

FATEMEH LAGHAEI

A dissertation submitted in partial fulfillment of the  
requirements for award of the degree of  
Master of Science (Information Technology – Management)

Faculty of Computer Science and Information Systems  
Universiti Teknologi Malaysia

JULY 2011

*To my beloved husband, parents, and siblings thank you  
for your understanding, encouraging and continuous support  
during my vital educational years.*

## ACKNOWLEDGEMENT

Praises to God for giving me the patience, strength and determination to go through and complete my study. I would like to express my appreciation to my supervisor, Associate Professor Dr. Othman bin Ibrahim, for his support and guidance during the course of this study and the writing of the thesis and Madam Lijah Rosdi for her friendly help.

Thanks also to my lecturers, Dr. Ab. Razak Che Hussin, Dr. Gholam Hassan Tabatabaei, Assoc.Prof.Dr. Rusni Daruis, and my friends for their kind assistance.

I would also like to thank all T-Systems Malaysia, Profitera Corporation Sdn. Bhd. and Arman Paya Communication personnel and representatives who have patiently participated in evaluation and testing.

I would also like to extend my thanks to my Husband who has given me the encouragement and support when I needed it. Finally, I would like to dedicate this thesis to my beloved husband and my parents. Without their support and trust I would have never come this far.

## ABSTRACT

Nowadays social networks become part of our life and using them for different purposes such as communities and business finds value. Using power of social network websites in business and enterprise become a dramatic value and made integration of existing Enterprise Information System (EIS) with social networks to use its data value a challenge and new requirement from enterprises. This study investigate on data and it's format of social networks beside type and structure of EIS in different level and integration methods to purpose a proper model for integrating them. Next, after comparing different integration strategy and integration technologies besides conducting questionnaire from few companies, Extract, Transform and Load (ETL) technology has been chosen as the main concept to propose Social Network Extract, Transform and Load (SNETL) model. This model can use to enhance an ETL framework for integrating existing EIS functionality and Social Networks in data repository level based on Enterprise requirement by integration project's team. Finally, this study used SNETL Model to enhance Talend Open Studio (TOS) as an ETL framework with developing an example Input and Output module to integrate Facebook news and the number of its comments and likes as a social network data Example and news submission in SugarCRM as an EIS example.

## ABSTRAK

Pada masa kini, rangkaian sosial menjadi sebahagian daripada kehidupan kita, dan ianya digunakan bagi tujuan yang tertentu seperti komuniti dan nilai terhadap perniagaan. Penggunaan kuasa laman rangkaian sosial dalam perniagaan dan enterprise akan menjadikan ianya bernilai dan luar biasa, di samping dapat mewujudkan integrasi di antara Sistem Maklumat Enterprise dan rangkaian sosial dengan menggunakan data yang bernilai tersebut sebagai satu cabaran dan keperluan baru bagi enterprise berkenaan. Kajian ini menjalankan penyiasatan ke atas data dan rekabentuk rangkaian sosial di samping jenis dan rangka kerja Sistem Maklumat Enterprise dalam aras yang berbeza untuk penghasilan model yang sesuai melalui kaedah integrasi. Seterusnya, hasil perbandingan strategi integrasi yang berlainan dan teknologi integrasi, serta soal selidik dari beberapa syarikat, maka teknologi Penghasilan, Penukaran dan Penyimpanan telah dipilih sebagai konsep utama bagi membangunkan Model Penghasilan, Penukaran dan Penyimpanan Rangkaian Sosial. Model ini boleh digunakan untuk meningkatkan keupayaan struktur Penghasilan, Penukaran dan Penyimpanan bagi mengintegrasikan fungsi Sistem Maklumat Enterprise yang sedia ada, dan penyimpanan data rangkaian sosial mengikut keperluan enterprise dengan mengintegrasikan pasukan projek yang terlibat. Kajian ini menggunakan Model Penghasilan, Penukaran dan Penyimpanan Rangkaian Sosial bagi penambahbaikan *Talend Open Studio (TOS)* sebagai rangka kerja Penghasilan, Penukaran dan Penyimpanan dengan membangunkan contoh modul input dan output bagi mengintegrasikan berita *Facebook* dan bilangan komen sebagai contoh data rangkaian sosial, dan penghantaran berita dalam *SugarCRM* adalah sebagai contoh Sistem Maklumat Enterprise.

## TABLE OF CONTENT

<b>CHAPTER</b>	<b>TITLE</b>	<b>PAGE</b>
	<b>TITLE</b>	<b>I</b>
	<b>DECLARATION</b>	<b>II</b>
	<b>DEDICATION</b>	<b>III</b>
	<b>ACKNOWLEDGEMENT</b>	<b>IV</b>
	<b>ABSTRACT</b>	<b>V</b>
	<b>ABSTRAK</b>	<b>VI</b>
	<b>TABLE OF CONTENT</b>	<b>VII</b>
	<b>LIST OF FIGURES</b>	<b>XI</b>
	<b>LIST OF TABLES</b>	<b>XIII</b>
	<b>LIST OF ABBREVIATIONS</b>	<b>XIV</b>
	<b>LIST OF APPENDIXES</b>	<b>XV</b>
<b>1</b>	<b>PROJECT OVERVIEW</b>	<b>1</b>
	1.1 Introduction	1
	1.2 Problem Background	2
	1.3 Problem Statement	3
	1.4 Project Objectives	4
	1.5 Project Scope	5
	1.6 The Project Importance	5
	1.7 Summary	6
<b>2</b>	<b>LITERATURE REVIEW</b>	<b>7</b>
	2.1 Introduction	7

2.2	Variety in Data Formats	8
2.2.1	Types of Heterogeneities in Data	9
2.2.2	Data Models Based on Data Formats	9
2.3	Information Systems	10
2.3.1	Types of Information Systems	11
2.3.2	Enterprise Information System	13
2.3.2.1	Customer Relationship Management	14
2.3.2.2	Social CRM	16
2.3.2.3	Data Warehouse	17
2.4	Concepts in Social Networks	19
2.4.1	Definition of Social Networks and Communities	19
2.4.2	Web Services	21
2.4.3	FOAF- Friend of a Friend	21
2.4.4	Google OpenSocial	22
2.4.5	OpenID 2.0	22
2.4.6	Issues and Challenges	24
2.4.6.1	Issues in Social Networks	24
2.4.6.2	Issues in Integrating Data in Social Networks	25
2.4.7	A Brief on a Famous Social Networks	26
2.4.7.1	Facebook	26
2.4.7.2	Graph API	27
2.5	Levels of Data Integration	28
2.6	Strategies of Data Integration	29
2.6.1	Global as View	30
2.6.2	Local as View	31
2.6.3	Peer-to-Peer Data Model (P2P)	32
2.7	Data Integration Technologies	32
2.7.1	Extracting, Transforming, Loading	32
2.7.1.1	Quick History of ETL	35
2.7.2	Enterprise Application Integration	35
2.7.3	Master Data Management and Customer Data Integration	37
2.7.4	Comparison of Data Integration Technologies	37

2.8	Summary	38
<b>3</b>	<b>METHODOLOGY</b>	<b>39</b>
3.1	Introduction	39
3.2	Project Methodology	40
3.3	Operational Framework	41
3.4	Phase Descriptions	42
3.5	Software Justification	44
3.5.1	Data Integration and Talend Open Studio	44
3.5.1.1	Operational Integration	45
3.5.1.2	Execution Monitoring	46
3.5.2	SugarCRM	47
3.6	Summary	48
<b>4</b>	<b>FINDING</b>	<b>49</b>
4.1	Introduction	49
4.2	The Value of Social Networks in Business	50
4.2.1	Social Networks Styles	52
4.2.1.1	Participating Business in Social Networks	53
4.3	Proposed Data Integration Technology for Enterprise Information Systems	54
4.4	Proposed Model Based on ETL for Enterprise Information Systems	57
4.5	Questionnaire Analysis	62
4.5.1	Criteria for Choosing Respondents	62
4.5.2	Analysis of the Questionnaire	63
4.5.3	Interview Analysis	64
4.6	Data Analysis	65
4.6.1	Part One	66
4.6.2	Part Two	67
4.6.3	Part Three	68
4.7	Testing SNETL Model	71
4.7.1	Facebook API	71
4.7.2	Install and Using SugarCRM	72



4.7.3	Enhance Talend Open Studio Framework Based on SNETL	73
4.7.4	Testing Product Publishing Integration	75
4.8	Interview Questionnaire Analysis	78
4.9	Summary	82
<b>5</b>	<b>DISCUSSION AND CONCLUSION</b>	<b>83</b>
5.1	Introduction	83
5.2	Achievement	84
5.3	Constraints and Challenges	85
5.4	Future Work	86
5.5	Aspiration	86
5.6	Summary	87
	<b>REFERENCES</b>	<b>88</b>
	<b>APPENDIX</b>	<b>93</b>

## LIST OF FIGURES

FIGURE NO.	TITLE	PAGE
2.1	Multiple profiles in social network websites	8
2.2	Four level pyramid model based on the different levels of hierarchy in the organization	12
2.3	Some examples of enterprise information systems	13
2.4	Analytical CRM	15
2.5	Diffrentiate between traditional CRM and social CRM	17
2.6	Map of Open ID	23
2.7	Global-as-view (GaV) model	30
2.8	Local-as-view (LaV) model	31
2.9	ETL diagram	33
3.1	The talend open studio interface	45
3.2	Operational integration	45
3.3	Execution monitoring	47
4.1	CRM companies integrate (Sparsely) with social networking platforms	51
4.2	Integrating social network's data by ETL processes	56
4.3	SNETL model	58
4.4	SNETL model, social network direction, load to social networks	59
4.5	SNETL model, EIS direction, extract from social networks	60
4.6	Education qualification of respondents	66
4.7	Interest of using EIS and social networks	68
4.8	Usage of social network	70

<b>4.9</b>	Levels of integration	70
<b>4.10</b>	Object data information from Facebook	72
<b>4.11</b>	Adding social network component to talend pallet	74
<b>4.12</b>	ETL job In talend open studio in social network direction	75
<b>4.13</b>	ETL job In talend open studio in EIS direction	77
<b>4.14</b>	SNETL page on Facebook after running ETL jobs	78
<b>4.15</b>	Answers of interview quetionnaire	80
<b>4.16</b>	Amount of benefit by using SNETL	81

**LIST OF TABLES**

<b>TABLE NO.</b>	<b>TITLE</b>	<b>PAGE</b>
2.1	The main information category on Facebook	27
2.2	Comparison of EAI and ETL	36
2.3	Comparison of data integration technologies	38
3.1	Phase descriptions	42
4.1	Question view	67
4.2	Answers of question one	69
4.3	Questions view	79

**LIST OF ABBREVIATIONS**

API	Application Programming Interface
BI	Business Intelligent
CDI	Customer Data Integration
CRM	Customer Relationship Management
DW	Data Warehouse
EAI	Enterprise Application Integration
EIS	Enterprise Information System
ERP	Enterprise Resource Planning
ETL	Extract, Transform, Load
GUI	Graphical User Interface
IS	Information System
IT	Information Technology
MDM	Master Data Management
OLTP	Online Transactional Process
SN	Social Network
SNETL	Social Network Extract, Transform and Load
SOA	Service Oriented Architecture
SOAP	Simple Object Access Protocol
TOS	Talend Open Studio
UDDI	Universal Description Discovery and Integration
UTM	Universiti Teknologi Malaysia
WSDL	Web Services Description Language
XML	Extensible Markup Language

**LIST OF APPENDICES**

<b>APPENDIX</b>	<b>TITLE</b>	<b>PAGE</b>
APPENDIX A	Questionnaire	93
APPENDIX B	Interview questionnaire	98
APPENDIX C	Facebook graph example	99
APPENDIX D	Facebook components for talend open studio	102

## **CHAPTER 1**

### **PROJECT OVERVIEW**

#### **1.1 Introduction**

Over the past few years social networks have become popular. They also have spread from websites to online playing games. Involvement and importance in social networks is becoming penetrating and data volumes rise dramatically; nevertheless, the lack of set up methodologies of the field of databases is in existing tools and models for data storage and manipulation in the area. For example, the formats which are ready for use need modification of collected data to their representational capabilities, often excluding potentially useful data, and also there are not any standard storage formats promoting reuse, sharing or combination of data sets from different sources.

Data integration is one of the aged investigation fields in the database area and has come out shortly after database systems were first introduced into the business world.

*“Data integration involves combining data residing in different sources and providing users with a unified view of these data [1].”*

To facilitate ease and performance of information retrieval is the main objective of data integration when queries execute on a large set of sources' data which cannot retrieve them one by one from sources in execution time such as the way in Service Oriented Architecture (SOA) model. The most common data integration approach in data layer is to provide a set of materialized views or summary tables over the different source to provide data inside database. This reason is one of main reason for integrating data in data layer by Extract, Transform and Load (ETL) model instead of integration on other layer.

## **1.2 Problem Background**

In today business and IT world, online communities and social networks have become the main concern for communication, fun and networking. The large number of online social networks in the different field has resulted in vast but varied information about people and their interesting. The rapid growth of social networking web sites in the workplace means companies can no longer ignore them. Furthermore, companies are trying to acquire access to the conversations on the social networks and participated in the dialogue. In order to use them with commercial view point to improve their service or business, we need to integrate them inside our organization process and Information Systems.

There are lots of useful data available in social networks that persuade the enterprise to import and combine its data with their product and service information



without upgrading their current systems or spending high cost. In order to reach these usages, we need to integrate Social Networks' data and our Enterprise in our organization repository with ability of updating both from each other. This is exactly the place where integrating the Enterprise Information Systems (EIS)'s and social networks in data layer becomes absolutely necessary. The current information inside companies database such as customer, products and services information's have a direct or indirect relation to pages, review and fan users in Social network that did not mapped or integrated together in most organization. They growth in separate line with extra effort and do not support each other to increase market, feedbacks, product quality and benefit improvement.

### **1.3 Problem Statement**

Today Business needs up-to-date information about products, services and market from different social such as social network for using in their business or services but we have different social networks around the internet and existing EIS which do not support this data exchanging. We need to integrate them with exist customer and product information in organization IS. Data integration is the problem of combining data residing at different sources, and providing the enterprise user with a unified view of these data. In real world of application the problem of data integration plays an important role. Besides, vast information is available on the social networks and this information is widely distributed and heterogeneous. Then, the main problems are:

- Different separated social networks for different requirement exist on the internet which define different requirement in integration for Enterprise.
- Enterprise needs transferring of required data around all social networks in or from their exist Enterprise Information Systems (EIS) such as Customer

Relationship Management (CRM), Enterprise resource Planning (ERP), Business Intelligent (BI) systems or Web Portals.

During conducting this project study, there are some important questions arise:

1. What are the values of social networks in EIS?
2. What are the technologies of data integration?
3. How EIS can integrate to social networks?

#### **1.4 Project Objectives**

The main objectives of this study are as follow:

1. To study and analyze data integration issues of social networking and enterprise database
2. To propose a model for integrating social network's data and enterprise information systems in data repository level.

## **1.5 Project Scope**

The scope of this study includes:

1. The study focuses on data integration of social network's data and EIS database.
2. The study evaluates and prototype the model by integrating an example of Facebook as a famous social network and SugarCRM as an EIS example.
3. The study is about integrating in data repository level and is focused on EIS database side.

## **1.6 The Project Importance**

Social networking refers to the online communities which are built based on common interests and activities. These communities at first were the place to make friends and meet romantic partners in which they have moved into the enterprise world, thus some companies are attempting to gain access to the data on social networks to raise awareness of their products or services and keep in touch with their existing or potential customers [2]. Integrating social networks' data and enterprise information systems can generate a lot of interest amongst the organizations because of the high potential usage attached to the development and popularity of Social networks besides organization Information Systems (IS) infrastructure.

## **1.7 Summary**

This chapter discussed the overview of the study which is brief introduction of data integration in social networks. Beside, the problem background of this study and also the main challenges to integrate data in social networks that mentioned in problem statement. There are two project objectives that need to successfully achieve in order to produce a model to integrate data.

## REFERENCES

1. Maurizio Lenzerini. *Data integration is harder than you thought*. Trento, Italy. September 5, 2001.
2. Deb Shinder. *Social Networking: Latest, Greatest Business Tool or Security Nightmare?.* 2009
3. Tim Finin, Li Ding, Lina Zhou and Anupam Joshi. *Social networking on the semantic web*, Tim Finin, Li Ding, Lina Zhou and Anupam Joshi.2005. The Learning Organization. vol. 5.
4. J.D. Ullman. Information integration using logical views. In F.N. Afrati and P. Kolaitis, editors. *Proceedings of the 6th International Conference on Database Theory (ICDT'97)*. January 1997. volume 1186 of Lecture Notes in Computer Science, pages 19–40, DelphiGreece,
5. Li Ding Tim Finin Anupam Joshi Rong Pan R. Scott Cost Yun Peng, Pavan Reddivari Vishal Doshi Joel Sachs. *Swoogle: A Search and Metadata Engine for the Semantic Web*. In proceedings of the 13th ACM Conference on Information and Knowledge Management , 2004.
6. Yutaka Matsuo and Masahiro Hamasaki and Yoshiyuki Nakamura, Takuichi Nishimura and Koiti Hasida. *Spinning Multiple Social Networks for SemanticWeb*.
7. Yutaka Matsuo, Junichiro Mori, Masahiro Hamasaki .*POLYPHONET: An Advanced Social Network Extraction System from the Web*.
8. 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.

9. Mark Hansen, Stuart Madnick, Michael Siegel. *Data Integration Using Web Services*.
10. Jayant Madhavan, Shawn R. Jeffery, Shirley Cohen, Xin (Luna) Dong, David Ko, Cong Yu, Alon Halevy. *Web-scale Data Integration: You can only afford to Pay As You Go*.
11. A.Y. Halevy. *Answering queries using views: A survey*. The VLDB Journal, 10(4):270–294, 2001.
12. 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.
13. Afraz Jaffri, Hugh Glaser and Ian Millard. *URI Identity Management for Semantic Web Data Integration and Linkage*.
14. David Recordon and Drummond Reed. *OpenID 2.0: A Platform for User-Centric Identity Management*.
15. Juliana Mitchell-Wong, Ryszard Kowalczyk, Albena Roshelova, Bruce Joy and Henry Tsai. *OpenSocial: From Social Networks to Social Ecosystem*.
16. A.Y. Levy, A. Rajaraman, and J.J. Ordille. *Querying heterogeneous information sources using source descriptions*. In Proceedings of the Twenty-second International Conference on Very Large Data Bases (VLDB'96), pages 251–262, Mumbai (Bombay), India, September 1996.
17. Okkyung Choi, Sangyong Han and Ajith Abraham. *Integration of Semantic Data Using a Novel Web Based Information Query System*.
18. Alon Y. HaLevy, Daniel S. Weld. *The Nimble XML Data Integration System, Denise Draper*.
19. Diego Calvanese, Elio Damaggio, Giuseppe De Giacomo, Maurizio Lenzerini and Riccardo Rosati. *Semantic Data Integration in P2P Systems*.
20. Fritz Lehman and Anthony G. Cohn. *The Egg/Yolk reliability hierarchy: Semantic Data Integration Using Sorts with Prototypes*.
21. Wayne Eckerson and Colin White. *Evaluating ETL and Data Integration Platforms*. The Data Warehousing Institute (TDWI), 2003
22. Li Xu and David W. Embley. *Automating schema mapping for data integration*. 2002. department of computer science, Brigham Young university.

23. 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.
24. Laudon, K.C. and Laudon, J.P. *Management Information Systems*, (2nd edition), Macmillan, 1988.
25. Bryan Bergeron. *Essentials of CRM: A guide to customer relationship management*. 2002, Wiley, USA
26. Jill Dyche. *The CRM handbook: A business guide to customer relationship management*. Addison Wesley. 2002
27. Brian Siu, Paul KM Kwan, Benedict lam, Peter de Vrise(editors). *Data mining, data warehousing & client/server databases*. proceedings of the 8<sup>th</sup> international database workshop (industrial volume).1997. Springer
28. Paul Greenberg. *CRM at the Speed of Light: Social CRM Strategies, Tools, and Techniques for Engaging Your Customers*. Fourth Edition. McGraw-Hill. 2009.
29. 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.
30. Reynol Junco and Jeanna Mastrodicasa .*The Net.Generation Generation: What Higher Education Professionals Need to Know about Today's Students*. (Washington, D.C.: NASPA, 2007).
31. Source: retrieved April 2010 from  
[http://www.crm2day.com/content/t6\\_librarynews\\_1.php?id=50036](http://www.crm2day.com/content/t6_librarynews_1.php?id=50036)
32. Franchise Tax Board's (FTB) Enterprise Architecture (EA) Data Center of Excellence. *Data Integration Strategy – v5.0*. Des 2008.
33. DestinationCRM.com (2009). *Who Owns the Social Customer?*, CRM magazine's in-depth report on the state of social media in CRM, June 2009.
34. Source: retrieve March 2011 form  
<http://statistics.allfacebook.com/pages>
35. Amanda Lenhart, Kristen Purcell, Aaron Smith, Kathryn Zickuhr. *Social Media and Young Adults*. Retrieved Feb 3, 2010 from  
<http://www.pewinternet.org/Reports/2010/Social-Media-and-Young-Adults/Part-3.aspx?view=all>

36. Source: retrieved March 2011 from <http://www.facebook.com/vspink#!/cocacola>
37. Source: retrieved March 2011 from [www.activerain.com](http://www.activerain.com)
38. Source: retrieved March 2011 from <http://developers.facebook.com/docs/reference/api/>
39. John Ward and Joe Peppard. *Strategic planning for information systems*. Third ed. 2002. Wiley.
40. Vrana. *How to approach the development of enterprise information system*. Czech University of Agriculture , Prague, Czech Republic.2004.
41. Rasik Phalak and Srinivasan Krishnan. *Data integration in social networks*. CSci 5707 – Principles of Database Systems. 2008.
42. R. Kimball, J. Caserta. *The Data Warehouse ETL Toolkit*. 2004: Wiley Publishing.
43. 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.
44. Denish Gupt. *The levels for successful enterprise business data integration process*, retrieved FEB 2011 form <http://ezinearticles.com/?The-Levels-for-Successful-Enterprise-Business-Data-Integration-Process&id=5855725>.
45. Jill Dyche, Evan Levy. *Customer Data Integration: Reaching single vesion of truth*. Wiley, New Jersey. 2006.
46. Source: retrieved June 2011 from <http://www.facebook.com/tsystemsmalaysia>