CSCE 2013 General Conference - *Congrès général 2013 de la SCGC*

Montréal, Québec
May 29 to June 1, 2013 / *29 mai au 1 juin 2013*

Data Acquisition Model for Capturing Labour Resources Data from Industrial Construction Projects

Ahmed Hammad, Ph.D., P.Eng., PMP, Project Controls Manager, In province Modules (IPM), Hebron Topsides Project, Atlantic Canada Business Unit, Eastern Canada Operations, WorleyParsons Canada, Suite 102, West Campus Hall, 240 Waterford Bridge Road, St.John's, Newfoundland, Canada A1E 1E2.

Simaan AbouRizk, Ph.D., P.Eng. Professor, Hole School of Construction Engineering and Management, Department of Civil and Environmental Engineering, University of Alberta, 3-014 Markin/CNRL Natural Resources Engineering Facility, Edmonton, Alberta, Canada T6G 2W2.

Yasser Mohamed, Ph.D., P.Eng. Assistant Professor, Hole School of Construction Engineering and Management, Department of Civil and Environmental Engineering, University of Alberta, 3-011 Markin/CNRL Natural Resources Engineering Facility, Edmonton, Alberta, Canada T6G 2W2.

Abstract: Proper management of labour resources is a key for success in any industrial construction project. A tremendous amount of labour data is generated and collected during planning and execution of these projects. Typically this data is not analysed nor transferred to useful knowledge that can be utilized for future projects. To address these problems, an integrated framework was developed based on a Knowledge Discovery in Data (KDD) model to enable the implementation of data warehousing and data mining techniques. The framework transfers existing multidimensional historical data from completed projects into useful knowledge for future projects. First, a synthesis of previous research is presented. Second, an inclusive analysis of the industrial construction domain is performed. Third, the concept of predefined progressable work packages is introduced to address issues in current data acquisition practices. This paper includes case studies of how the model was implemented to capture data from three very different, mega-multi-billion-dollar, global projects.

1 INTRODUCTION

Industrial construction projects involve a large number of stakeholders with different, sometimes conflicting, interests. Stakeholders include owners, Project Management Team (PMT) engineers, suppliers, fabricators, constructors, environmental and other governmental agencies, plant operators and maintainers and the general public. Nearly all industrial construction projects are performed as a set of smaller projects, each of which is performed by a contractor. These contractors include Engineering, Procurement and Construction Management (EPCM) offices, fabrication shops and module assembly yards. These contractors are utilizing labour resources to provide their services and rely solely on a continuous supply of projects to generate their revenues. Contractors manage multiple projects in a changing environment using the same pool of resources. According to Huemann et al. (2007), the number and the size of the projects are constantly changing in this environment. The supply of projects is dependent on market conditions, which are difficult to predict. This dependency causes significant uncertainty and makes it very difficult to estimate the required amount of hours to complete the expected projects (workload) and the necessary resources to perform this workload (capacity). While managing labour resources, a large amount of data is generated, collected, and stored in different formats. Construction contractors need a reliable method to capture this data in a format that enables them to utilize new techniques such as data mining for better prediction of their workload and the resource level required to manage that workload.

The objective of this paper is to illustrate the development of a data acquisition model as a component of an integrated framework for managing labour resources data in a multiple-project environment utilizing the concepts of Knowledge Discovery in Data (KDD), data warehousing, and data mining. The Cios et al. (2007) hybrid model was modified, as illustrated in Figure 1, and adapted to achieve this objective. The model starts by understanding the problem domain, after that analyzing the problem data and then developing the data acquisition model. The acquisition model prepares the data for mining and produces dynamic On Line Analytical Processing (OLAP) reports and graphs.

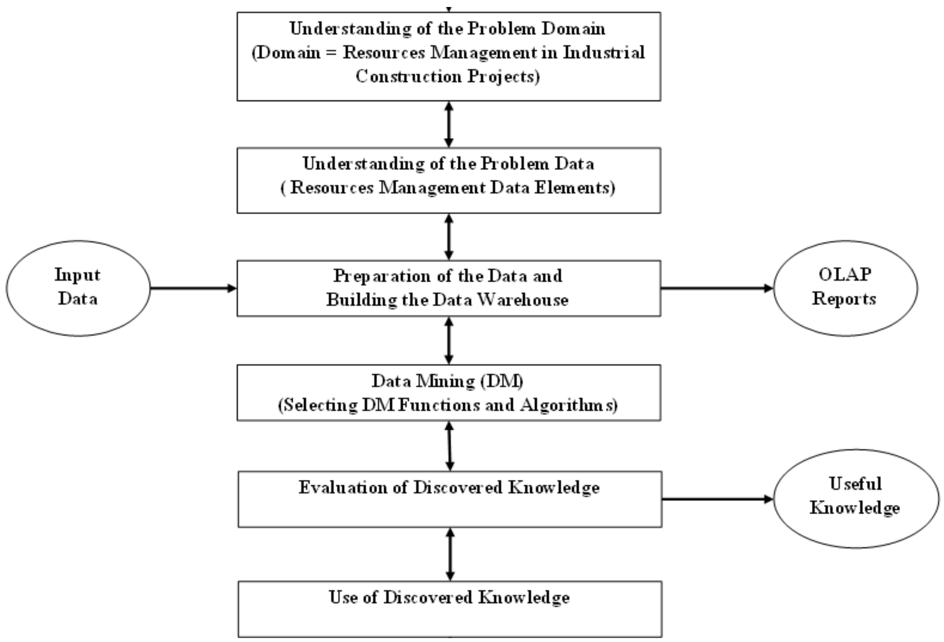

Figure 1: The modified hybrid KDD model

2 LITERATURE REVIEW

Traditional planning techniques such as Critical Path Method (CPM) assume unlimited availability of resources, which is not realistic. Resource leveling techniques (time-constrained scheduling) assume an unconstrained amount of resources in order to maintain the schedule; meanwhile, resource allocation techniques (resource-constrained scheduling) assume constrained availability of resources to modify the schedule accordingly. In commercially available software applications, the two terms are used interchangeably.

Several techniques are used for predicting resource requirements for projects. These techniques can be grouped into three main categories: heuristic rules, numerical optimization, and genetic algorithms. Heuristic approaches primarily utilize pre-defined rules to find an acceptable solution to a problem (Fendley 1968, Dumond 1992, Lu and Li 2003). Numerical optimization utilizes mathematical modeling to predict the optimum project level (Mohanty and Siddiq 1989, El-Rayes and Moselhi 2001, Vaziri et al. 2007). A Genetic Algorithm is a computing technique that imitates the real-life evolution process in order to find approximate solutions to optimization problems. The procedure starts with generating a set of random solutions as single strings called a chromosome. These solutions go through a cycle of generation, evaluation, selection and recombination based on "*survival of the fittest*" until the termination condition is reached (Chan et al. 1996, Hegazy 1999, Kandil and El-Rayes 2006).

Data warehousing, On Line Analytical Processing (OLAP) and data mining were recently introduced to resource management research. Data warehouses are dedicated, read-only, and non-volatile databases that centrally store validated, multidimensional, historical data from Operation Support Systems (OSS) to be used by Decision Support Systems (DSS). OLAP techniques are used to enable dynamic analysis of the warehouse data according to the decision-makers' needs (Codd 1993). OLAP techniques include: roll-up and drill-down, slice and dice, and data pivoting (Fan 2007). Han et al. (2006) defined data mining as "the analysis of observational datasets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owners." Data mining techniques can be either supervised or unsupervised and can be grouped into four categories: Clustering, Finding Association Rules, Classification, and Outlier analysis (Cios et al. 2007).

Chau et al. (2002) combined the concepts of data warehousing, Decision Support Systems (DSS) and OLAP to develop the Construction Management Decision Support System (CMDSS). Rujirayanyong and Shi (2006) developed a Project-oriented Data Warehouse (PDW) for contractors. Fan et al. (2008) used the Auto Regression Tree (ATR) data mining technique to predict the residual value of construction equipment. This technique represents an easily interpreted non-linear regression model.

3 UNDERSTANDING THE DOMAIN OF INDUSTRIAL CONSTRUCTION PROJECTS

Industrial construction projects can be on-shore or off-shore, green field or brown field. They typically require several levels of multi-discipline engineering, rolling-wave planning, Value Engineering (VE), Hazardous and Operability Analysis (HAZOP), Constructability Operability and Maintainability reviews. O'Neill (1989) classifies industrial construction projects as *Team* projects to emphasize the role of multiple teams of highly-skilled individuals in successful completion of this type of project. People or human capital represents the number one area of focus in industrial project management accompanied with processes and technologies (Badiru 2008). Figure 2 represents the typical labour resources status in contractors who are involved in managing multiple industrial construction projects simultaneously.

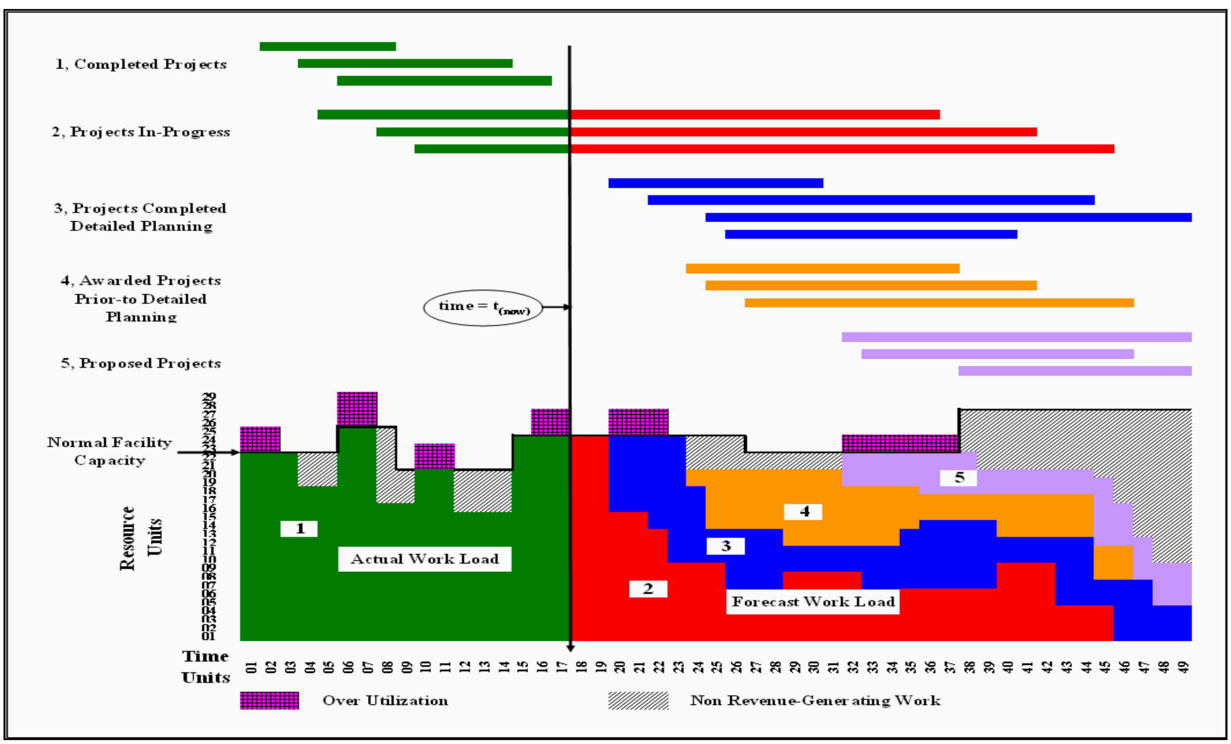

Figure 2: Typical labour resources status in a multiple-project environment

All workload prior to time = $t_{(now)}$ is actual workload, which is the aggregation of work hours from completed projects and the completed portions of in-progress projects. All workload after time = $t_{(now)}$ is expected workload and it is composed of the forecast of required work hours per time-unit for all production resources from the incomplete portion of in-progress projects, projects that have detailed plans, awarded projects and proposed projects. All contractors try different methods to predict their optimum human resources level to meet their expected workload to minimize resources overutilization, idleness and turnover of staff.

To develop the data acquisition model, a set of predefined elements and objects were developed to fully represent the industrial construction domain. First, project phases and sub-phases are clearly defined. These phases are: Pre-Engineering, Engineering, Procurement, Construction and Commissioning and Start-up (C&SU). The engineering phase is broken down into: Front End Engineering and Design (FEED), Detailed Engineering and Design, Shop Drawings, Procurement Support, Construction Support and As-Builting. The procurement phase is broken down into: Engineering Support, Requisition, Bidding and Awarding (RBA), Subcontract Administration and Materials Management. The construction phase is broken down into: Engineering Support, Fabrication, Assembly, and Site Installation. The Commissioning and Start-up (C&SU) phase is broken down into: Engineering Support (ES), Pre-Commissioning and Commissioning. Resource utilization profiles are captured for each sub-phase to be used for future estimating of resource requirements.

Second, a set of labour industrial resources, grouped in a hierarchical Resources Breakdown structure (RBS), is defined. The hierarchy consists of four levels: Group, Department, Discipline, and Individual. The labour resources were categorized in three groups Production, Project, and Support Services. Production services has 3 groups: Engineering, Procurement and Construction. For engineering services, process, mechanical, piping, HVAC, civil, structural, architectural, electrical, instrumentation and controls departments were defined. For each department, the management, coordination, engineering, design and administration disciplines are utilized.

Third, a predefined set of production packages is developed for collecting historical labour resources data. Process Flow Diagrams (PFD's), Piping and Instrumentation Diagrams (P&ID's), Single Line Diagram (SLD's), Line Designation Tables (LDT's) are examples of these production packages for engineering services. Work Packages are used as the building block for capturing the data in the data warehouse. The concept of work packages is highly utilized in industrial construction, but in a format that’s not consistent between projects, and not suitable for data mining. As shown in Figure 3, work packages are used to allocate all the necessary attributes to the collected data to enable efficient data mining. The data related to resource utilizations represented as weekly planned, earned and spent hours is collected at the package level using the predefined set of resources and other data attributes (metadata).

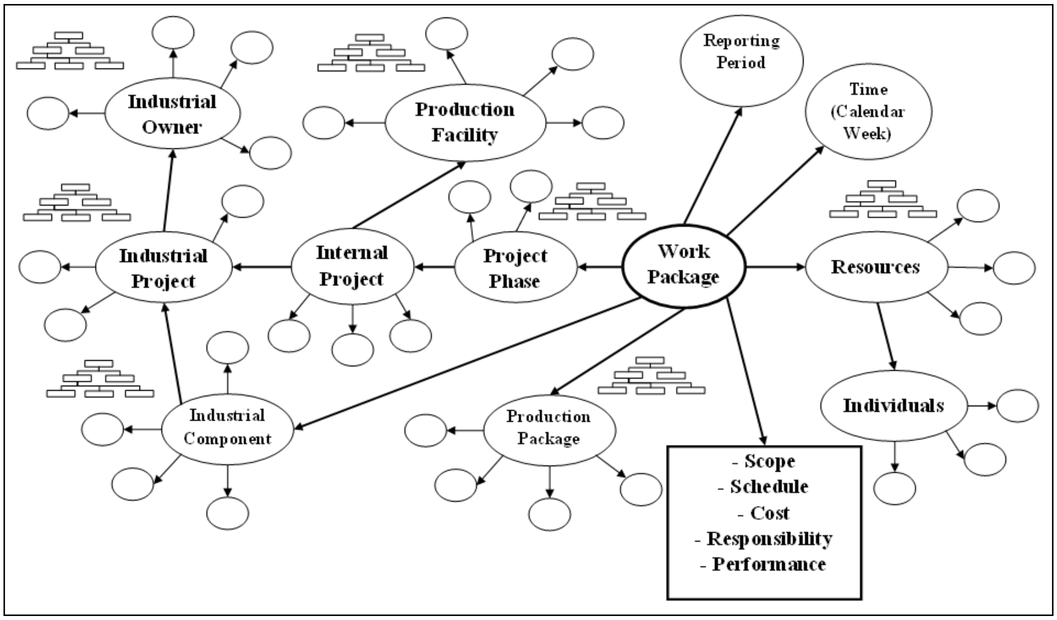

Figure 3: Using work packages to capture labour resources data

4 DATA COLLECTION AND BUILDING THE DATA WAREHOUSE

The engineering labour resources data was collected from 3 different multi-billion dollar projects to verify the applicability of the data acquisition model. The first project is a complete oil sand development that has ore preparation, froth treatment and extraction plants. The second project is an LNG plant with 3 gas dehydration and compression trains. The third project is an offshore oil drilling, production and storage platform. A data warehouse was developed in MS Access utilizing the snowflake schema, as shown in Figure 4.

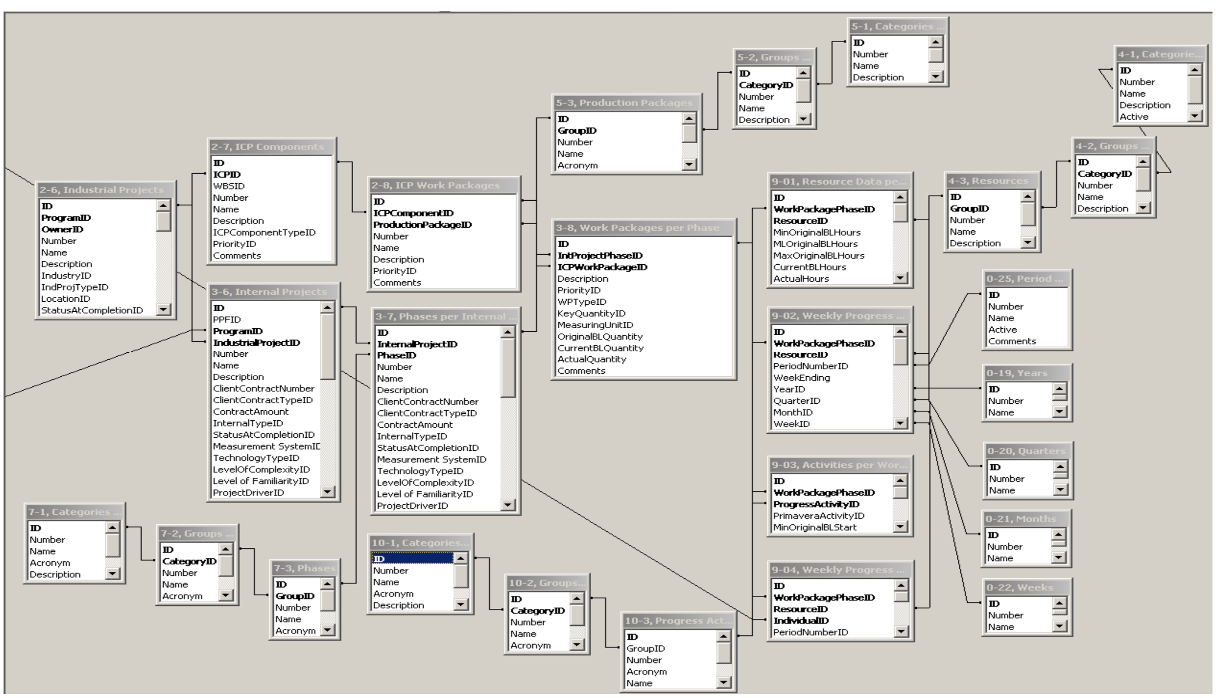

Figure 4: The work package presented as a snowflake schema

A user interface (backend) and an OLAP output module (frontend) was also developed. Examples of the dynamic OLAP reports and graphs are shown in Figure 5-a shows an example of a report showing the performance of a single resource in multiple projects. Figure 5-b shows a summary report of resource utilization over time periods using the pivoting technique. Figure 5-c demonstrates an example of the dynamic graphical output showing a comparison of resource utilization over four years.

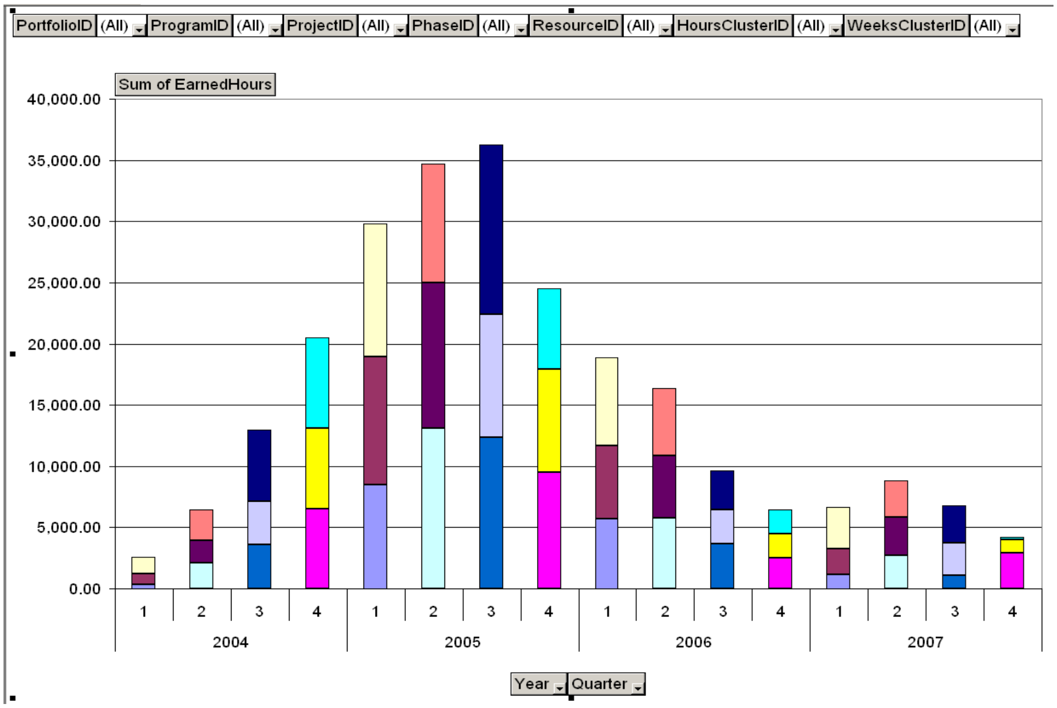

Figure 5-a

Portfolio	(All)																		
Program	(All)																		
ProjectI	(All)																		
					2004 Total	2005												2005 Total	2006
Phase	Resource	Data	11	12		1	2	3	4	5	6	7	8	9	10	11	12		
6	1	Sum of CurrentBLHour	10.38	33.54	**893.69**	64.98	14.71	6.00	1.39									**87.07**	
		Sum of EarnedHours	10.11	9.52	**571.19**	39.42	14.66	9.63	0.19									**63.90**	
		Sum of ActualHours	12.75	6.50	**583.50**	36.25	15.25	6.50	1.00									**59.00**	
	2	Sum of CurrentBLHour	176.20	377.83	**2,459.51**	191.23	240.96	190.34	37.52	80.71	66.34	57.40	1.21	56.08	97.97	129.55	326.82	**1,476.13**	215.2
		Sum of EarnedHours	160.93	222.50	**1,518.91**	149.13	193.58	83.22	29.19	47.76	54.01	55.07	1.66	39.35	67.70	84.32	97.21	**902.19**	145.6
		Sum of ActualHours	192.00	216.00	**1,619.00**	127.00	169.25	92.00	30.50	55.00	60.50	68.00	1.00	28.25	64.00	65.75	133.75	**895.00**	124.5
	4	Sum of CurrentBLHour	221.85	277.61	**3,316.34**	125.61	24.76	29.12	29.50	22.08	74.45	15.09	27.00	27.00	104.88	79.97	15.73	**575.19**	119.5
		Sum of EarnedHours	226.35	173.27	**1,779.73**	80.04	20.18	30.05	20.88	10.75	32.38	13.34	24.98	23.05	132.37	89.36	15.39	**492.76**	102.2
		Sum of ActualHours	200.00	136.50	**1,729.75**	98.75	34.50	44.00	17.50	11.25	25.50	19.00	25.00	37.25	124.00	84.00	17.75	**538.50**	102.2
	5	Sum of CurrentBLHour			**206.48**							62.98	58.68		19.41			**141.07**	
		Sum of EarnedHours			**145.89**							78.69	46.01		6.74			**131.44**	
		Sum of ActualHours			**194.75**							49.00	29.75		7.50			**86.25**	
	6	Sum of CurrentBLHour	29.05	20.04	**1,282.90**	35.61	132.76	272.62	134.42	209.43	95.92	175.91	155.45	135.84	122.12	222.32	301.13	**1,993.54**	103.3
		Sum of EarnedHours	17.11	8.48	**831.01**	37.60	102.29	95.15	101.57	174.95	85.82	91.70	99.58	101.71	88.44	104.21	146.69	**1,229.70**	64.8
		Sum of ActualHours	41.50	12.50	**828.00**	58.50	129.00	103.50	108.25	160.50	135.00	85.50	100.50	98.00	132.25	126.00	163.00	**1,400.00**	55.2
6 Sum of CurrentBLHours			**437.48**	**709.02**	**8,158.93**	**417.43**	**413.19**	**498.09**	**202.82**	**312.22**	**236.72**	**311.38**	**242.34**	**218.92**	**344.38**	**431.84**	**643.68**	**4,273.00**	**438.1**
6 Sum of EarnedHours			**414.51**	**413.77**	**4,846.73**	**306.20**	**330.70**	**218.04**	**151.82**	**233.46**	**172.21**	**238.80**	**172.23**	**164.10**	**295.25**	**277.88**	**259.30**	**2,819.99**	**312.7**
6 Sum of ActualHours			**446.25**	**371.50**	**4,955.00**	**320.50**	**348.00**	**246.00**	**157.25**	**226.75**	**221.00**	**221.50**	**156.25**	**163.50**	**327.75**	**275.75**	**314.50**	**2,978.75**	**282.0**
7	1	Sum of CurrentBLHour	736.87	1,055.84	**9,476.21**	1,030.18	300.64	680.79	758.95	51.89	29.76	176.14	219.24	420.94	5.09	151.48	234.09	**4,059.19**	161.0
		Sum of EarnedHours	306.83	362.16	**5,612.86**	434.14	239.92	457.64	402.58	63.61	9.01	159.30	111.34	191.60	9.42	126.03	199.22	**2,403.81**	182.0
		Sum of ActualHours	372.75	341.00	**5,518.50**	359.00	259.25	320.75	404.50	44.25	14.25	176.25	107.25	200.75	10.50	112.00	195.75	**2,204.50**	151.0
	2	Sum of CurrentBLHour			**446.31**			41.70	109.75	241.30	65.21	97.44	177.97	196.80	180.71	352.83	477.19	**1,940.90**	572.1
		Sum of EarnedHours			**470.94**			21.51	109.22	109.78	92.43	83.69	127.17	107.70	171.18	160.02	337.34	**1,320.04**	243.9
		Sum of ActualHours			**541.50**			22.75	85.00	113.75	94.50	81.00	98.25	168.75	168.25	148.50	277.75	**1,258.50**	231.5
	4	Sum of CurrentBLHour	13.73		**35.16**	139.16	9.42				119.23	249.56	62.47	72.46	33.66	7.24	237.71	**930.91**	264.9
		Sum of EarnedHours	10.80		**15.67**	53.96	10.30				76.32	53.25	71.03	74.16	52.11	3.26	133.52	**527.91**	179.1
		Sum of ActualHours	18.75		**30.50**	37.00	12.50				55.00	42.50	59.75	87.25	52.25	3.50	139.75	**489.50**	152.2
	5	Sum of CurrentBLHour											16.86	123.66	128.60	59.95	145.26	**474.32**	84.2
		Sum of EarnedHours											12.57	66.57	29.38	27.80	130.66	**266.97**	130.1
		Sum of ActualHours											16.50	45.50	19.50	38.50	114.25	**234.25**	78.2
	6	Sum of CurrentBLHour	29.67	6.25	**560.06**							56.37	70.03	142.55	189.04	38.06	110.67	**606.72**	229.3
		Sum of EarnedHours	27.05	10.24	**269.66**							73.88	50.08	67.80	61.87	19.98	96.91	**370.51**	260.5
		Sum of ActualHours	36.25	8.25	**271.25**							90.25	61.25	45.50	42.50	39.00	103.50	**382.00**	187.7
7 Sum of CurrentBLHours			**780.26**	**1,062.09**	**10,517.74**	**1,169.35**	**310.05**	**722.49**	**868.69**	**293.19**	**214.21**	**579.50**	**546.57**	**956.42**	**537.09**	**609.56**	**1,204.91**	**8,012.03**	**1,311.7**
7 Sum of EarnedHours			**344.68**	**372.39**	**6,369.13**	**488.10**	**250.22**	**479.14**	**511.80**	**173.39**	**177.77**	**370.11**	**372.19**	**507.83**	**323.97**	**337.09**	**897.64**	**4,889.24**	**995.8**
7 Sum of ActualHours			**427.75**	**349.25**	**6,361.75**	**396.00**	**271.75**	**343.50**	**489.50**	**158.00**	**163.75**	**390.00**	**343.00**	**547.75**	**293.00**	**341.50**	**831.00**	**4,568.75**	**800.7**
8	1	Sum of CurrentBLHour	416.23	425.81	**3,264.10**	416.29	284.09	165.05	202.75	1,442.52	475.92	1,891.49	459.21	6.37	1.12			**5,344.81**	
		Sum of EarnedHours	387.93	144.59	**1,845.24**	285.08	270.80	156.47	142.78	828.21	667.51	773.65	278.83	7.08	2.03			**3,412.44**	
		Sum of ActualHours	314.00	130.50	**1,574.50**	287.50	241.00	140.75	162.50	624.75	751.50	806.00	363.50	12.00	1.75			**3,391.25**	
	2	Sum of CurrentBLHour	397.85	42.66	**2,213.37**	13.53			37.92	113.23	63.52	57.79	26.67	156.74	307.33	266.59	70.38	**1,113.71**	
		Sum of EarnedHours	381.04	58.27	**1,857.49**	9.09			25.42	73.59	29.56	56.88	16.93	130.30	152.79	136.22	17.39	**648.18**	
		Sum of ActualHours	332.50	54.00	**1,733.75**	20.25			31.25	56.25	37.00	61.75	16.75	145.75	154.75	92.00	19.50	**635.25**	
	4	Sum of CurrentBLHour	190.23	140.23	**1,093.71**	42.11								44.68	25.91	61.99	81.36	**256.03**	47.4
		Sum of EarnedHours	228.99	126.19	**974.21**	21.02								47.29	22.82	49.73	36.31	**177.17**	11.3
		Sum of ActualHours	254.50	86.50	**1,049.50**	17.25								43.75	25.50	30.75	32.00	**149.25**	15.2
	5	Sum of CurrentBLHour	130.56		**1,255.69**	40.47	20.14	1.74	10.71	24.08	2.53	5.01		95.43	36.96	96.24	78.28	**411.58**	145.6
		Sum of EarnedHours	145.17		**871.92**	20.13	5.86	1.19	8.78	21.41	1.39	3.81		27.04	55.76	100.56	65.49	**311.40**	93.5
		Sum of ActualHours	121.25		**911.25**	25.25	17.75	2.00	8.50	45.50	1.50	4.50		32.25	44.00	108.00	91.75	**381.00**	66.0

Figure 5-b

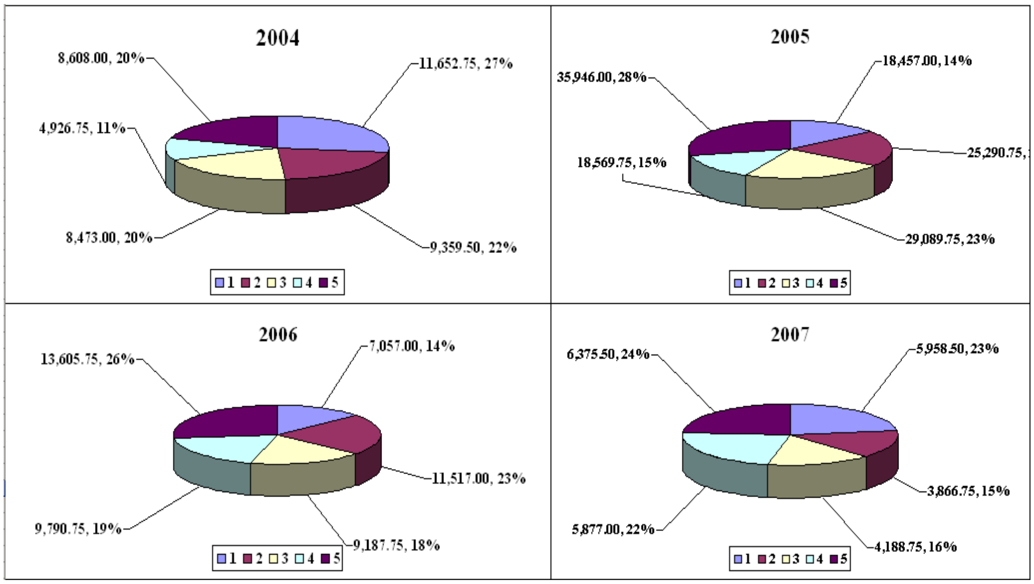

Figure 5-c

Figure 5a-c: Examples of the dynamic OLAP reports and graphs

5 CONCLUSION

Although projects are temporary and unique (PMI 2008) they can be easily broken down to smaller work packages that share similar attributes amongst multiple projects. Using these work packages to capture labour resources data and store it in a data warehouse ready for data mining provides a wealth of knowledge to industrial contractors and academic researchers. In current practices, most of the collected data from completed projects is stored without being analyzed and the available knowledge is neither captured nor utilized for better estimating of resource requirements in new projects.

This paper presented the data-acquisition model of an integrated framework for managing labour resources data in multiple industrial construction projects. The framework is built on a KDD model to transfer the collected multidimensional historical data from completed projects to useful knowledge for new projects. The developed framework was tested in 3 very different mega projects to validate its ability to efficiently store labour resources data in the required format. Implementing data mining techniques to the collected data is presented in another paper.

Acknowledgements

The authors would like to thank the NSERC Alberta Industrial Research Chair in Construction Engineering and Management and WorlyParsons Canada for their support of this research.

References

Badiru, A. B., Badiru, A., and Badiru, A. 2008. *Industrial project management: concepts, tools, and techniques.* CRC Press, Boca Raton.

Chan, W., Chua, D., and Kannan, D. 1996. Construction Resource Scheduling with Genetic Algorithms. *Journal of Construction Engineering and Management,* 122(2): 125.

Chau, K. W., Cao, Y., Anson, M., and Zhang, J. 2002. Application of data warehouse and decision support system in construction management. *Autom.Constr.,* 12(2): 213-224.

Cios, K. J. 2007. *Data mining: a knowledge discovery approach.* Springer, New York.

Codd, E. F., Codd, S. B., and Salley, C. T. 1993. *Providing OLAP (On-line Analytical Processing) to user-analysts: an IT mandate.* Codd and Date, San Jose, CA, USA.

Dumond, J. 1992. In a multi-resource environment, how much is enough? *Int J Prod Res,* 30(2): 395-410.

El-Rayes, K., and Moselhi, O. 2001. Optimizing resource utilization for repetitive construction projects. *J.Constr.Eng.Manage.,* 127(1): 18-27.

Fan, H. 2007. Leveraging operational data for intelligent decision support in construction equipment management. PhD thesis, University of Alberta (Canada), Canada.

Fan, H., AbouRizk, S., Kim, H., and Zaiane, O. 2008. Assessing residual value of heavy construction equipment using predictive data mining model. *J.Comput.Civ.Eng.,* 22(3): 181-191.

Fendley, L. G. 1968. Toward development of complete multiproject scheduling system. *Industrial Engineering,* 19(10): 505-515.

Han, J., and Kamber, M. 2006. *Data mining: concepts and techniques.* Morgan Kaufmann; Elsevier Science distributor, San Francisco, Calif; Oxford.

Hegazy, T. 1999. Optimization of resource allocation and leveling using genetic algorithms. *J.Constr.Eng.Manage.,* 125(3): 167-175.

Huemann, M., Keegan, A., and Turner, J. R. 2007. Human resource management in the project-oriented company: A review. *International Journal of Project Management,* 25(3): 315-323.

Kandil, A., and El-Rayes, K. 2006. Parallel genetic algorithms for optimizing resource utilization in large-scale construction projects. *J.Constr.Eng.Manage.,* 132(5): 491-498.

Lu, M., and Li, H. 2003. Resource-activity critical-path method for construction planning. *J.Constr.Eng.Manage.,* 129(4): 412-420.

Mohanty, R. P., and Siddiq, M. K. 1989. Multiple projects — Multiple resources constrained scheduling: A multiobjective analysis. *Engineering Costs and Production Economics,* 18(1): 83-92.

O'Neill, J. J. 1989. *Management of industrial construction projects.* Nichols Pub., New York.

Project Management Institute (PMI) 2008. Project Management – Body of Knowledge (PM-BoK).

Rujirayanyong, T., and Shi, J. J. 2006. A project-oriented data warehouse for construction. *Autom.Constr.,* 15(6): 800-807.

Vaziri, K., Carr, P. G., and Nozick, L. K. 2007. Project planning for construction under uncertainty with limited resources. *J.Constr.Eng.Manage.,* 133(4): 268-276.