Workflow logs Analysis System for Enterprise Performance Measurement

Jin Lai, Yushun Fan
Department of Automation, Tsinghua University, Beijing 100084, China
laijin00@mails.tsinghua.edu.cn fan@cims.tsinghua.edu.cn

[Abstracts]

Workflow logs that record the execution of business processes offer very valuable data resource for real-time enterprise performance measurement. In this paper, we proposed a novel scheme that uses the technology of data warehouse and OLAP to explore workflow logs and create complex analysis reports for enterprise performance measurement. Three key points of this scheme were studied: 1) the measure set; 2) the open and flexible architecture for workflow logs analysis system; 3) the data models in WFMS and data warehouse. We also provide a case study that shows the validity of our scheme.

[Keywords]

Workflow; business performance measure; data warehouse; OLAP; workflow logs analysis

1. Introduction

"If we cannot measure our business, we cannot manage it." With its unique importance for business control and improvement, enterprise performance measurement has arisen enormous interests all over the world. Because of its direct reflection of an organization's ongoing operations, business process measurement acts an important role in many performance measurement frameworks such as the notable balanced scorecard, the EFQM's businecess excellence model, performance prism and activity-based costing. Through business process analysis and measurement we can evaluate the enterprise's operating efficiency and even economic profit.

New commerce age puts forward new requirements on business performance measurement system. Gone are the days when managers are willing to wait until a couple of weeks for their performance reports, now they want them online and real-time. In this part, we begin to think of constructing business performance measurement system based on WFMS (workflow management system), a system that defines and automates business processes which are widely used in many enterpises. When executing business processes, WFMS documents all the steps of a business process, so workflow logs contain the information about each action taken in it, for example at what time which actor use which resource to perform which task in which business process. This information is very valuable for real-time business performance measurement in that 1) it describes the actual execution of the concurrent workflows in different aspects, such as time, human, resource, processes etc.; 2) it is real-time, the data are updated automatically; 3) the data model in workflow management system provide very good structure (business – business unit – business process – activities) for business performance measurement, which fits perfectly well with the prevailing performance measurement frameworks. For these reasons, **workflow**

management system (WFMS) can be and should be an important source for real-time enterprise performance measurement.

In this paper we proposed a data warehouse based multi-dimensional workflow logs analysis scheme that can support enterprise performance measurement. The rest of this paper is organized as follows: in Section 2, we will have a brief overview of the related work about workflow log analysis. In Section 3 we will discuss the measurement framework for workflow logs analysis. In section 4, the architecture of the data warehouse based workflow logs analysis systems are addressed. In section 5, we will discuss the multi-dimensional data model in workflow log data warehouse. In section 6, a case study is present. Finally in section 7, we draw some conclusions.

2. Related work about workflow logs analysis

Workflow logs analysis is a new and active research area in the recent years. There is quite some research work in this part. W.M.P van der Aalst[2] proposed a machine leaning based method to discover new workflow models from workflow logs. Sen'ichi Onoda [3] also proposed a method to refine the parameters of a workflow model through analyzing its execution logs, and then use simulation method to analysis the system's bottle-net.

And Eder proposed an idea of adapting data warehouse and OLAP for workflow logs analysis.[1] Their work did give us some inspiration, but it still has some drawbacks, for example: 1) they proposed this scheme just to answer the complex queries for system manager, while didn't take systematic consideration on enterprise performance measurement, so they failed to propose the measurement framework required by business performance measurement; 2) They didn't discuss the architecture to construct an open and flexible workflow log analysis system; 3) their data model for workflow logs analysis is based on their own workflow meta-model and thus is not general enough and cannot be used by other WFMSs.

To solve these problems we propose a more comprehensive sheme and studied its three key points: 1) the measurement framework for business performance measurement; 2) the architecture of workflow logs analysis system; 3) the common data model in WFMS and the multidimensional data model in workflow log data warehouse that considers more elements than Eder's.

3. Measurement framework of workflow log analysis

To support enterprise performance measurement, what should we analyze from workflow logs? In this part we proposed its measurement framework which defines the analysis objects, analysis measures and the requirement of measure reports based on the specifications of the prevailing business performance measurement frameworks and the information available in WFMS. As is shown in Figure1, the analysis objects include such elements as organization (the employees and the cooperative organizations outside), resoruce, business process, activity and the business strategy behind them. And the Analysis measures can be devided into three sets: time measure set, quality/ quality measure set and the cost measure set, among which time measure set is the largest. Time measures are used to analyze the efficiency of the enterprise system such as the competence of a certain employee or vendor, the efficientcy of a process, the utility of a type of resource which are very important to locate the bottleneck. Time measure set include many measures, such as Execution Duration, Estimated Execute Duration, Waiting Duration for

Activation, Waiting Duration for Resource, Waiting Duration, Resource occupancy Duration, Overtime, Surplus Time, their respecitve average value and some ratio value among them. And productivity and quality measures here refer to the execucation times of processes/activities and their failed ratio. Cost measure is an option measure set here, although there is no cost information in most WFMSs, if integrated with other information system such as PDM, ERP etc., we can fill the cost measure and realize activity based costing. The analysis report have three 4 perspectives, in which the manager can analysis the measures with the focus of a specific time duation, a process/acitivity model, a role or a person in an enterprise's organization sturcture or a type of resource in different abstract level.

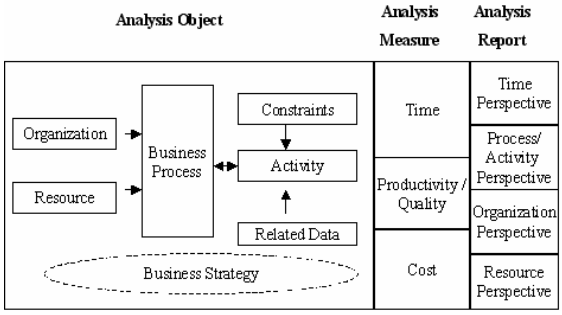

Figure 1 Measure framework for workflow log analysis

4. System architecture for workflow logs analysis

The architecture of workflow logs analysis system should support such functions as: 1) the information stored in data warehouse is updated real-time from workflow management system; 2) the system not only offer reports for performance measurement but also support the users to make customized query; 3) the system is open for integrating other data resource outside WFMS and other measurement modules to extend its function. According to these requirements, we proposed an architecture that is shown in the figure 2.

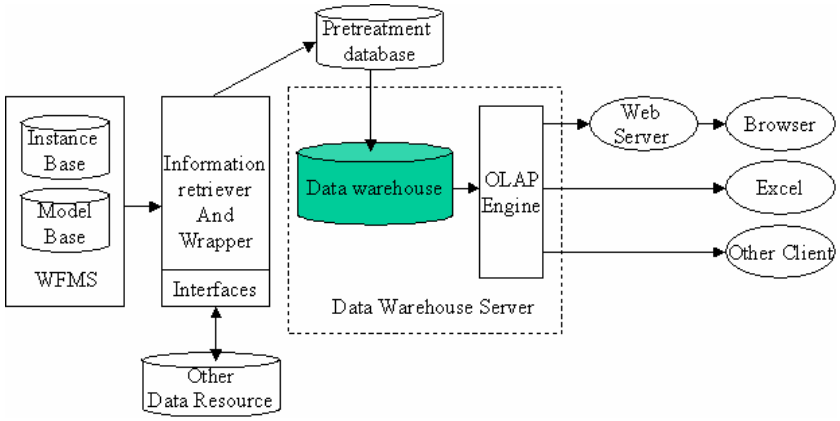

Figure 2 Architecture of the workflow logs Analysis system

WFMS logs are stored in the instance base and model base of WFMS. The information retriever and Wrapper module takes charge of collecting the data in WFMS in real-time, making necessary format transformation, and then attaching them to the pretreatment database of workflow log data warehouse. We proposed pretreatment database as the direct source of data warehouse server while not the databases in WFMS, because: 1) the data store in WFMS is too trivial, for example: it is impossible for it to offer factor table that is crucial for data warehouse; 2) some data formats in WFMS do not fit the requirement of data warehouse, for example: the time

format used in WFMS cannot be calculated in data warehouse; 3) some data are lack inWFMS, for example: time stamp. In this case, pretreatment database is put forward to solve these problems. Beside workflow logs, other data resource can be accessed through the Data retriever and wrapper model. Based on the data stored in pretreatment database, data warehouse server will construct cubes, a kind of multi-dimension data model, for workflow log analysis. OLAP engine support complex operations on the data warehouse, such as drill down, roll up, slicing, dicing, etc., which allow analysts to rapidly derive reports in different levels of abstractions and view the data from different perspectives. Here we adopt the Microsoft's SQL 2000 analysis server as our data warehouse server. The client applications access the mulit-dimensional data in warehouse through the interfaces of OLAP engine.

5. Cube model in workflow log data warehouse

The multi-dimension data model in workflow log data warehouse is very important for our system, which should not only specify the elements in workflow models and their complex relations but also meet the requirement of the measurement framewok that is constructed for enterprise performance measurement. Although our system is developed based on CimFlow, a workflow management system developed by CIMS Center, Tsinghua University, China. We adopt the meta-data model framework defined by WFMC[12] (Workflow management coalition), so that the system will be more easy to integrate data from other WFMSs. For the detail of its meta model, please refer to [12].

The data model is shown in figure 3. In this cube model, there are six dimensions: Workflow model unit (activity and process) dimension, Workflow instance dimension, Organization dimension, Resources dimension, Time dimension and Measures dimension. The time dimension is a kind of simple dimension, which has abstract levels of year, quarter, month, week and day. The workflow model unit dimension is a kind of Parent-Son dimension that includes the elements of processes and activities. In this dimension the elements are structured as a tree: a process may includes some activities and sub-processes. Workflow instance is put forward as an independently dimension, because the workflow model dimension is Parent-son dimension, it cannot support further amalgamation under its nodes in tree structure. The organization dimension has four different hierarchies, which are respectively role-human, human-role, department-human and human-department. With different hierarchies, we can view data through different abstract level. The resource dimension is a kind of snow dimension, who has the hierarchy of resource type – resource. In this part, the applications are included in resource dimension as a special resource type, "application". The measure dimension contents not only the all the measures defined in the previous section but also the raw data to calculate such measures. The raw measures are: 1) Start Time; 2) Activate time; 3) Finish Time; 4) Maximum Estimated Finish Time; 5) Minimum Estimated finish Time; 6) Apply resource Time; 7) Get resource Time; 8) Return Resource Time; 9) Task successful or not. Other measures are calculated based on the raw measures, so they are named as calculated members. Thus this data warehouse schema contains information about the build time properties as well as the run time properties of the workflows. The build time properties are given in the dimension structures and the run time properties are represented by the data values.

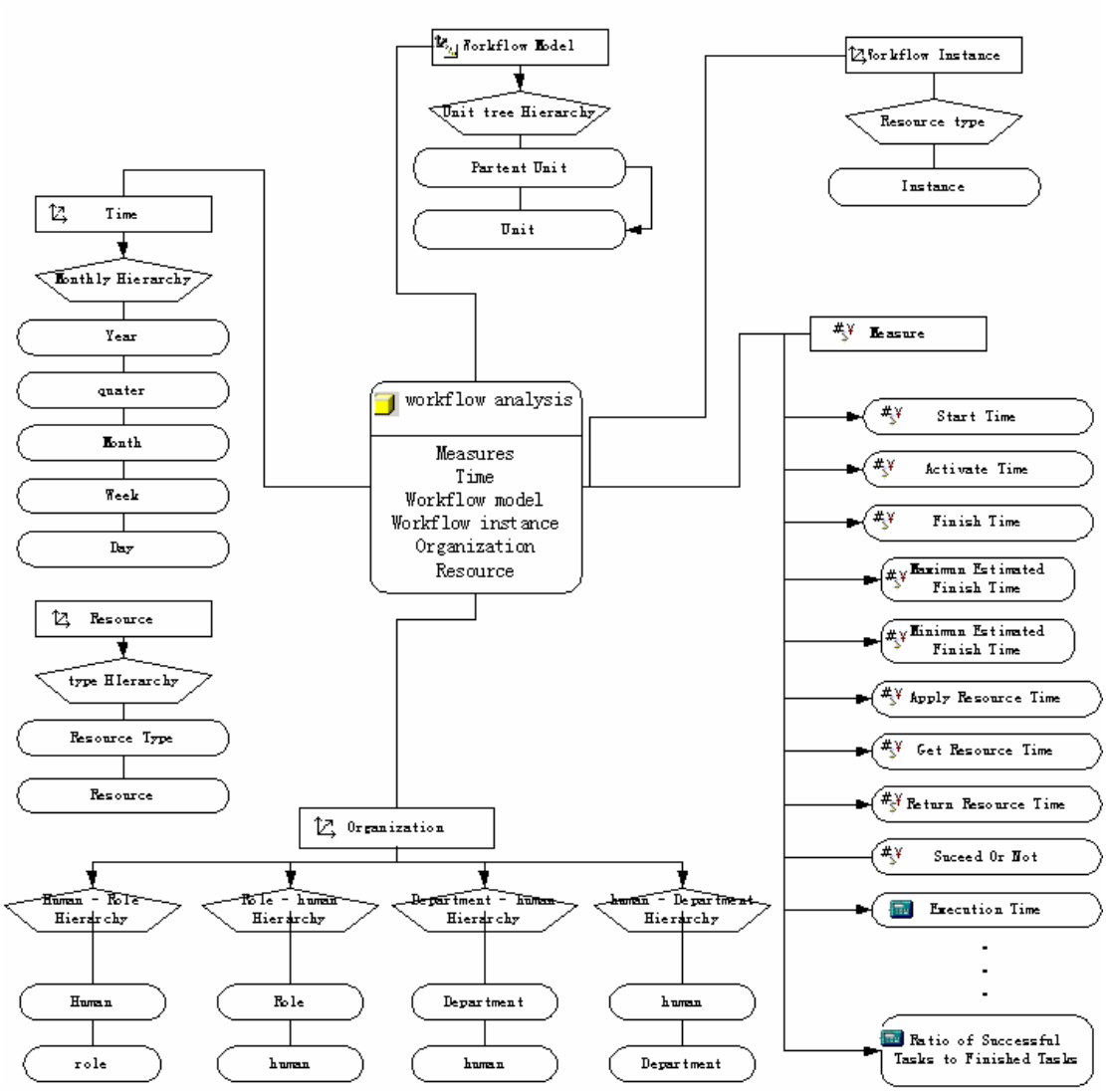

Figure 3 the Cube model in data warehouse

6. Case study

The goal of the case study is prototypical implementation of a data warehouse based workflow logs analysis system to analyze the validity of the theoretically elaborated concepts.

As is mentioned, the prototype adopts CimFlow as the source WFMS and use the instance database which stores 4, 0000 instance activities executed in the time period form Jan. 2000 to Dec. 2000, which is relative enough to make multi-dimension workflow logs analysis. Users can use browser or Excel to read the analysis report or make customized analysis with the graphical interfaces. The browser user interface is shown in the following figure 4.

Figure 4 User interface for workflow logs analysis system

Figure 4 is the advanced analysis page. In this page, users can choose to make analysis in different perspectives of time, organization, resource or instance. Figure 4 is a time perspective page. In this page, firstly users should specify the value of each peripheral dimension of this perspective, for example in time perspective, workflow model dimension, organization dimension, and resource dimension, then roll down data from year to any abstract level of time dimension (they can even roll down to from time dimension to instance dimension) and then chose the measure items from measure lists. Correspondingly the system will create chart diagram for analyzing, including the charts and data forms.

Despite of the customized query, the system also offers some predefined reports according to the managers' requirements. Figure 5 is the report on the information of the top 20 tasks that are finished ahead of time, which include the instance name, task name, time, actor and some measures. The reports can be updated real-time which are important for business control.

Figure 5 A report on the top 20 tasks that are finished ahead of time

7. Conclusion

Enterprise performance measure is a very active and promising field in the world. In such a global economy, there is a strong demand on real-time business performance. In this case, workflow logs that record the execution of business processes offer very valuable data resource for enterprise performance measurement. In this paper we proposed a novel scheme that uses the technology of data warehouse and OLAP to explore workflow logs for enterprise performance measurement. Firstly based on the prevailing enterprise performance measurement frameworks and the information available in workflow management system (WFMS), we proposed the measurement framework for workflow logs analysis. Then we put forward an open and flexible architecture for our workflow logs analysis system that not only support effective real-time multi-dimensional data analysis but also offer interfaces to integrate other data source or other measure modules to extend the function of our system. We also study the key point of our system, the multi-dimension data model in workflow log data warehouse. Finally in the case study, we study the functionality of our system, which can not only support its users to analyze the real-time performance of the resource, human, activities and processes in different perspective and different abstract level but also allow them view the measures in a long period of time, which is very valuable for business control and other decision making.

Reference

[1] Johann Eder, Georg E. Olivotto and Wolfang Gruber. A data Warehouse for Workflow Logs. EDCIS 2002, LNCS 2480, pp. 1-15, 2002.

[2] W.M.P. van der Aalst and B.F. van Dongen. Discovering Workflow Performance Models form Timed Logs. EDCIS 2002. LNCS 2480, pp. 45-63, 2002-4-20

[3] Sen'ichi Onoda, Masaki Yumoto, Tetsuya Maruta, Norithisa Komoda, Takashi Kobayashi. Bottleneck Detection Analysis for workflow Improvement. IEEE. 1997.

[4] Amit Basu and Robert Blanning. Metagraph Transformations and workflow Analysis. IEEE. 1997.

[5] Andy Neely. Business Performance measurement: theory and practice. Cambridge University Press 2002.

[6] Huiping Lin, Yushun Fan. Research on the workflow based enterprise measurement framework. Computer Integrated Manufacturing System, Vol 7. NO. 12 2001 pp. 47-32

[7] Ren Mnglu, Zju Weidong, Yag shanlin. Data Oriented Analysis of Workfow Optimization. Proceeding of the 3 rd World Congress on Intelligent Control and Automation. June28-July2, 2000.

[8] Neely, A.D. Performance Measurement: Why, what and How.. Lodon: Economist Books.1998.

[9] WorkGroup 1. Glossary: A workflow management coalition specification. Workflow Management Coalition, Brussels, Belgium, V 1.1 Final , 1994

[10] Inmon W.ll, Building the data warehouse, 2nd edition, John Wiler Computer Publishing, 1996.

[11] Angela Bonifati, Fabio Casati, Umeshwar Dayal and Ming Chien Shan. Warehousing workfklow data: challenges and opportunities. In the VLDB Journal, pages 649-652.

[12] Workflow Management Coalition. http:// http://www.wfmc.org/information/info.htm