AN APPROACH TO OPTIMIZE DATA PROCESSING IN BUSINESS PROCESSES

CSE 718

Professor: Michalis Petropoulos
Megha Ramesh Kumar
meghar@buffalo.edu

OVERVIEW

In this paper, they have proposed a framework for the optimization of data processing in business processes in order to optimize business process revenues and profits. A multi-stage control strategy is suggested for this optimization task. A rule based optimization of business processes is introduced using rewrite rules and a control strategy to transform the business process into a more efficient one with respect to data management. This paper focuses its study on business process execution language [BPEL].

The two programming layers, function layer and the choreography layer, have been explained in detail. The function layer consists of executable components in the form of Web Services that carry out basic activities. The choreography layer specifies a process model defining the execution order of activities. An example based on an order handling scenario has been explained. In this example, a detailed optimization of SQL activities at each step reduces the number of times the database has to be queried. The optimization leads to a reduction of data volume transferred between the database level and the workflow processing level. The overhead that is associated with the processing of each SQL statement is saved because only a single statement is left. These optimization steps results in a remarkable performance improvement independent from the underlying database.

The main contributions to the paper are:

- It demonstrates the opportunities, generality and benefits of the main approach by means of a case study.
- There is a multi-stage control strategy for comprehensive treatment of the task on optimization.
- The query optimization has been extended to a level of complex business tasks which are expressed as business processes.
- There is an insight on data flow and control flow of a business process for reasoning upon optimization decisions.

The paper explains the control strategy involved. The condition part of the rewrite rules is the part needed to preserve the semantics of the process. The action part defines the transformations applied to the process provided the corresponding condition is fulfilled. The following slides then goes into the classification of rewrite rules. They are classified into Activity merging rules, Updating merging rules and Tuple-to-Set Rules. The control strategy exploits the dependencies among the rewrite rules. It divides the overall process in several optimization spheres and applies rewrite rules considering their dependencies. An algorithm called the optimization sphere is

explained. Here the optimization spheres which are the parts in the process where application rules can be applied. This algorithm traverses the sphere and applies control strategy, applies the rule without changing the semantics of the process. After certain conditions are met the rule is applied.

DETAILED COMMENTS

The paper shows a promising approach to extend the optimization of business processes to data processing affairs. The paper however does not go too much into the technical details. It gives an overview of how this approach would be useful in the industry. Too many assumptions have been made in this paper making the optimization steps in practical applications a lot more complex than it seems. This paper focuses on keeping things simple making it easy for the user to understand. The paper provides a proper explanation reason while commenting on the disadvantages of the current approaches. This paper also provides justification which is proper backed by results obtained while proposing the new approach.

This paper is the first to combine several concepts from the programming language and database communities and that presents an integrated approach.

In analogy to federated database systems which integrate data sources and provide a homogeneous database schema for heterogeneous sources, this framework contains an optimizer, complementing the query optimizers of the database management systems that are responsible for executing the data management activities of a process.

Coarse-grained optimization is an approach for optimizing sequences of data processing tasks. Similar to CGO, this optimization approach is based on rewrite rules.

There was a question asked on the example presented in the paper. While combining the steps in the example, what if data dependency was not there between the activities of the for loop? In this case the Web Service invocation cannot be merged into a SQL Activity and hence no further optimization is possible.