Using Temporal Databases for Managing Business Process Execution Data

Master Thesis Proposal

for obtaining the academic degree

Diplom-Ingenieur/in

Software Engineering & Internet Computing

submitted by

Diana Toskova
Matriculation number 1428256

Department
Informatics Faculty, Vienna University of Technology

Advisor
Ao. Univ. Prof. Dipl. -Inf Dr. Ing Jurgen Dorn

Wien, 15.12.2017

(Submitter Signature)

(Advisor Signature)

1. Problem Statement

Processes surround us everywhere - from simple activities, performed in our everyday life, to complex business processes in organisations, involving many actors and elaborate actions. No matter of the type of organisation (educational, social, governmental, business, etc.), it is concerned with a various types of processes [7]. Typical process involves a number of actions (or simple tasks), different actors, which may be human actors or some computer systems/machines, various decision points and at the end some (or many) outcome(s) [1]. The Business Process Management (BPM) deals with such processes, involving different techniques for managing and monitoring processes, in order to enable process re execution, automation and optimisation. It is essential for every organisation to find a way to optimise its processes, in order to increase the quality, time to market and effective delivery of their products. Improvement goals mostly include reduction of costs, production time, errors and personnel involved, but also aim at better consistency and adaptability of processes and making them easily automatable and repeatable. The processes and the actors, involved in them, should be able to adjust themselves easily to new requirements and unexpected events. The general idea is to make the processes and the organisation itself more flexible [1]. A software that stores data from instances of process executions and allows execution of different queries can be beneficial for managers of organisations as they can easily monitor their processes and come up with ideas for improving them. The problem with traditional relational databases is that they do not offer (good) support for temporal data (which often exists in process execution entries).

Almost any organisation has to deal with temporal data while managing their processes. Many of the nowadays applications, such as: finance, insurance, law, medical, academic, etc. include temporal aspects. The data often includes some starting or ending time points of processes or single actions in them, or some information that will be valid only in a specific interval of time. Such applications should benefit of integrated temporal data support in the DBMS. Current relational databases store information, which is true at the current time. In contrast, temporal databases provide the possibility of storing information, which is valid in specific time point (interval, etc.). This time can be in the present, but also in the past or in the future. Such temporal databases provide the possibility to store and manage temporal information. They provide additional data types, related to time - instants, intervals and periods [2]. Additionally, the temporal databases introduce three types of times - current, sequenced and non sequenced. Temporal primary keys and constraints and automatic

dealing with time intervals when updating or deleting records enrich the possibilities for dealing with time-related instances. Furthermore the temporal databases provide different kinds of time-oriented statements.

The aim of the thesis is to answer the following research questions:

- What are the typical queries, executed on process execution data? Do they often include temporal aspects?
- Can the use of temporal databases and temporal queries be beneficial for querying process execution data in terms of time efficiency or simplicity?

2. Expected Results

This thesis should explore the additional opportunities for dealing with temporal data that temporal databases provide and investigate their benefits for the process management. The general goal is to evaluate, if the use of such temporal databases can be beneficial for operating on process execution data and if it can somehow contribute to improving process management. It should answer the research questions from the previous section by firstly exploring the literature sources regarding the topics.

The first step in the practical part of the thesis will be to model a scenario for an example process. Using an existing process execution engine, the process will be executed a sufficient number of times (for ex. 1000), and the execution instances data will be stored in a database. A goal of the thesis is the creation of prototype that will be used for evaluating the additional benefits of the use of temporal databases.

The thesis should examine what the typical database queries executed on process data could be and to test, if the additional capabilities of temporal databases could be beneficial for executing such queries. It should examine the time efficiency of temporal queries (if possible) and compare their simplicity with the traditional database queries. The results will be evaluated using the developed prototype. Another goal of this thesis is to come up with a methodology for managing relevant temporal data.

3. Methodology and Approach

The results from the previous section will be achieved using the following methodological approach:

- Literature research - literature will provide a comprehensive overview of the topics of Business Process Management, Workflow Modelling, Temporal Databases (in comparison with traditional relational databases), etc..
- Designing a process scenario - a scenario for some example process will be created using existing workflow and business process modelling tools. As part of the thesis should be chosen an appropriate modeling tool. The scenario should be representative for processes, in which temporal data is present.
- Process execution - after exploring different process execution engines, one will be used for executing the example process.
- Developing prototype - A simple software will be implemented, which will store sufficient amount of the process instances (for ex. 1000) using existing relational/temporal database.
- Query execution - Some queries, regarding temporal aspects of the data, will be executed.
- Evaluation of results - The the example process scenario will be used to evaluate the prototype. The traditional relational queries will be compared with the temporal relational queries. It should be evaluated, if the additional capabilities (temporal operators, transformation rules, query optimisation rules, additional time-related data types, etc) simplify the querying of process execution data.

4. State of the Art

The Business Process Management arises in the early 90s with the trend of organising business around processes. The basic idea behind it is that each product or service, delivered to the market, is the result of many actions and decisions, taken by different actors. Each process takes some input, performs some tasks on it and some output is expected, which should bring value to the customer. For the business is extremely important to understand the interactions between those activities, in order to improve the corresponding processes. Basic characteristic of successful company is the ability to adapt existing

products to the market demands and to promptly create new functionalities and products, when needed [3].

The interests of the business administration community from the one side include reducing costs of products/services and increasing customer satisfaction. On the other side lies the computer science point of view, where scientists try to investigate the structural characteristics of processes, using abstractions of real business processes [3]. Software engineers are also interested in bettering the existing and creating improved software systems, that can support the business process management. Especially in the current century of extreme development of computers in general, Internet of Things and Artificial Intelligence, the information systems play even more important role in the area, since they support more and more the processes at companies. At the beginning of process management, information technology supported it by performing some processes, then it continued by contributing to the improvement of processes [8].

The Business Process Model and Notation (BPMN) is widely used for specifying business processes as models. The so called business process management systems are used to coordinate everything that is involved in a process. Weske defines such system as "a generic software system that is driven by explicit process representations to coordinate the enactment of business processes" [3]. Famous BPM systems today is for example Bpm'online, which gives companies the possibility to control and improve their processes and to adapt them to the market demands. Another BPM solutions include Nintex, Clarizen, Camunda BPM, TIBCO BPM, KiSSFLOW, aMember, Zoho Creator, Wrike, Process Street, etc.. Many of those already run on the cloud and also give the user easy access through mobile device [4].

Some of the current relational databases provide support for temporal data, which includes dealing with queries on time intervals. Temporal databases on the other hand provide additional temporal data models and query languages improvement, that can provide a considerable simplification of dealing with temporal data [6]. Some of the queries appear to be extremely difficult to formulate using SQL and considerably more simple when using temporal queries. The traditional DBMS concentrate on supporting the representation of temporal data – such as intervals, but lack on improvements on the relational database engine [6].

The temporal databases define the so called temporal tables. Temporal tables are also called system-versioned tables and provide information about the entries in the table at any point in time, compared with the traditional relational databases that provide information, that is valid only at the current time [5]. This allows querying updated or deleted rows of data, unlike the normal tables, which allow only querying the current data. Each temporal table has two period columns (start and end column), used by the system, in order to manage a period of validity for each row of data. It also contains a reference to a history table, used to store previous versions of rows (modified or deleted). Temporal databases use simple schemas and make querying such data much simpler. The application code is simpler and understandable, which automatically makes it more maintainable and easier to ensure its correctness.

The temporal data feature was introduced in ANSI SQL 2011 and some support is existing in the current database management systems. For example with the SQL Server 2016 version, support for temporal tables was provided as build-in feature. The temporal feature is under current development, which means that currently there are some limitations [5].

5. Relevance to the Curricula of Software Engineering & Internet Computing

The thesis will combine the fields of workflow and business process management and temporal databases. The aim is to test the possible benefits of such temporal databases for the BPM and especially for process improvement. The most relevant courses are the following:

188.924 Workflow Modelling and Process Management

188.910 Advanced Software Engineering

188.483 Knowledge Management

6. References

[1] Dumas, Marlon, La Rosa, Marcello, Mendling, Jan. and Reijers, Hajo A.. Fundamentals of Business Process Management. 1st ed. Berlin: Springer, 2013

[2] Snodgrass, Richard. *Developing Time-Oriented Database Applications in SQL*. San Francisco, California: Morgan Kaufmann Publishers, 2000

[3] Weske, Mathias. *Business Process Management. Concepts. Languages. Architectures.* 2nd ed. Berlin, Heidelberg: Springer, 2012

[4] *Business Process Management Software*. capterra.com. N.p., 2017. Web. Accessed 11 Dec. 2017

[5] *Temporal Tables*. microsoft.com. N.p., 2016. Web. Accessed 10 Dec. 2017

[6] Dignös, Anton, Böhlen, Michael Hanspeter, Gamper, Johann, Jensen, Christian S. *Extending the kernel of a relational dbms with comprehensive support for sequenced temporal queries*. ACM Transactions on Database Systems, 41(4):1-46. Zurich Open Repository and Archive, University of Zurich, 2016

[7] Zimanyi, Esteban. *Temporal Databases*. Department of Computer & Decision Engineering (CoDE). Universit e Libre de Bruxelles, 2015

[8] Tick, Jozsef. *Workflow Model Representation Concepts*. Institute for Software Engineering, John von Neumann Faculty, Budapest, 2006