FAKULTÄT FÜR INFORMATIK
Faculty of Informatics

Bestätigung über die Abfassung der Diplomarbeit

Name der Betreuerin/des Betreuers: Ao.Univ.Prof. Dipl.-Inf. Dr.-Ing. Jürgen Dorn

Name der/des Studierenden: Thomas Sonnleitner, Bsc, Bsc

Die Betreuerin/Der Betreuer und die Verfasserin/der Verfasser der Diplomarbeit bestätigen mit ihrer Unterschrift, dass Sie die Richtlinien der Fakultät für Informatik zur Verfassung einer Diplomarbeit sowie zur Erstellung des Diplomarbeitsposters (siehe: http://www.informatik.tuwien.ac.at/dekanat/abschluss-master) zur Kenntnis genommen haben und darüber informiert wurden, dass deren Nicht-Einhaltung die Abweisung der Diplomarbeit bzw. des Posters zur Folge haben kann.
Die Betreuerin/Der Betreuer und die Verfasserin/der Verfasser nehmen weiters zur Kenntnis, dass eine Sperre der Diplomarbeit nur in besonders begründeten Fällen und nur im Ausmaß von max. einem Jahr genehmigt werden kann (Die Sperre muss im eigenen Interesse der/des Studierenden liegen und ist vor einer etwaigen vertraglichen Geheimhaltungsverpflichtung durch den/die Studiendekan/in genehmigen zu lassen).

Unterschrift der Betreuerin/des Betreuers

Unterschrift der/des Studierenden

FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

Persistence of workflow control data in temporal databases

DIPLOMA THESIS PROPOSAL

to write a diploma thesis for obatining the academic degree of

Diplom-Ingenieur

in

Business Informatics

by

Thomas Sonnleitner, Bsc, Bsc
Registration Number 0751500

to the Faculty of Informatics

at the TU Wien

Advisor: Ao.Univ.Prof. Dipl.-Inf. Dr.-Ing. Jürgen Dorn

Vienna, 11th April, 2019

Thomas Sonnleitner	Jürgen Dorn

Contents

CHAPTER 1

Motivation and problem statement

Every organization has to manage a number of processes of different kinds such as order, purchase, issue-resolutions or other internal processes. A process contains activities, tasks, events, decision points, a number of actors as well as physical or immaterial objects and leads to one or several outcome(s). It might be rather simple or more complex.
The way of how a process is designed and executed affects the quality of service which is perceived by customers, employees or other involved persons. Therefore, organizations do have interest to set up their processes in the most effective way. A process analysis founds the basis for processes improvement (or redesign). "Typical example of improvement objectives include reducing costs, reducing execution times and reducing error rates" [DRMR13].

The possibility to analyze but also monitor and execute processes more efficiently motivates organizations to digitalize their processes, or at least support process execution with IT-systems. This happens e.g. in the financial services industry. Financial institutions expect ongoing digitalization to result not only in lower costs but also in faster execution times. On the one hand, there is a need to because of regulatory requirements on reporting timeliness such as demanded by IASB financial reporting standards IFRS 9 [Int14]. On the other hand, also clients demand fast processing times of loan applications, account openings or other services, as they are already used to in other (online) services.

A digitalized process management is supported by a business process or Workflow Management System (WfMS), which governs or automatizes the process flow, supports the approval management and furthermore tracks and monitors process execution. Digitalization of processes and the availability of process metadata (e.g. information about process executions) enables organization to analyze and improve processes.

However, conventional database systems, in which process metadata is usually stored, do not adequately support the handling of time-related data. In particular, querying data with regard to time-related aspects can be complex and associated with long run-times during execution. Temporal databases focus on temporal aspects in data and promise easier querying and faster execution times. Therefore, they offer various kinds of time oriented statements, specialized data types and a more applicable data organization[Sno00].

To get better insights into the benefit of using such temporal databases, this work will deal with the following research questions:

1. What are typical queries executed on process execution data during process analysis, considering temporal aspects?

2. Can the use of temporal databases and temporal queries be beneficial for querying time-related data in terms of time efficiency or simplicity in context of business process management?

CHAPTER 2

Expected results

To improve a process, the current process and thus its process execution data needs to be analyzed. In a first step, this thesis will determine possible information needs which arise during process analysis and their according database queries on process execution data. The result will be a list of typical queries which are executed during business process analysis.

Furthermore, the thesis will identify features of temporal databases which improve the handling of workflow related data, being relevant for process management. Thus, extended query features, but also data types and other characteristics of temporal databases will be reflected in context of process management. Additionally, dimensions that could represent such an improvement will be selected and measures defined, e.g. dimension time efficiency with measure execution time of queries.

The thesis will perform an evaluation of a temporal database in comparison to a conventional databases in context of process management, considering previously defined benchmark dimensions. Therefore, a sample process as well as sample queries on process execution data will be defined. A prototype will be developed which generates mock-up process execution data, performs sample query executions and captures measures of benchmark dimensions. The thesis will document and interpret captured measures to subsequently determine if and how temporal databases be beneficial for process management.

CHAPTER 3

Methodological approach

The methodological approach consists of the following steps:

1. **Literature Review**
   Literature provides the theoretical background in the field of process and workflow management, including approaches and measures for quantitative process analysis and notations for process modelling. Furthermore, literature will be reviewed regarding features and capabilities of temporal databases. Additionally, already existing approaches on how to compare different database technologies and measure performance of query executions will be screened as basis for determination of benchmark dimensions and measures.

2. **Design of a sample process**
   A sample process will be designed using BPMN 2.0 in order to generated sample process execution.

3. **Creation of sample process execution**
   Sample process execution of various process runs will be generated with the help of a workflow engine. The created data should be representative for Business Process Management (BPM) related data sets with focus on temporal data.
   The process execution will be stored using two different database management systems: PostgreSQL as a conventional relational database system and a research prototype by the University of Bolzano as temporal database system.

4. **Definition and execution of queries on process execution**
   Based on literature review, a variety of different queries on process execution will be defined. These queries should cover the whole range of different kinds of data requirements for process analysis. The sample process should also be used as scenario to show, how process execution can be used to find process optimization potential.

5. **Evaluation of query execution performance**
   After the defined queries have been executed on a conventional and a temporal database system, the performance of these database systems in handling the query requests will be evaluated in the defined dimensions with the defined measures.

CHAPTER 4

State-of-the-art

In the 90s, processes and process design moved into the focus of companies for the first time. In the period before, management's attention was primarily dedicated to the functional organization of companies. As this lead to inefficient and bureaucratic way of work, the management's attention moved from the distribution of responsibilities and tasks along a hierarchical organization to the design of more efficient processes.
As not only processes have been re-designed, but also organizations it-self became process-oriented, BPM arose. "BPM provides concepts, methods, techniques and tools that cover all aspects of managing a process - plan, organize, monitor, control - as well as its actual execution"[DRMR13]. So BPM is not only about planning and organizing processes, but also about managing a company process-centered.
Thereby, BPM can be seen as continuous cycle, including the phases process identification, process discovery, process analysis, process redesign, process implementation as well as process monitoring and controlling as described by [DRMR13].

In phase process discovery, identified processes are documented in form of as-is process models. In comparison to textual descriptions of processes, a graphical model allows stakeholders to more easily comprehend a process and avoids misinterpretations originating in the ambiguity of text [DRMR13]. One of the most well known graphical notations for business process modelling is Business Process Model and Notation (BPMN) in version 2[Wes12].

BPM phase process analysis is particularly relevant to this thesis as performance measures of a process are quantified based on process execution data. This is done in order to determine how well a process is performing with respect to its performance objectives. Typical performance measures are e.g. processing time and waiting time but also quantifications of the dimensions quality, cost and flexibility[DRMR13]. Not only the calculation of performance measures, but also other process analysis techniques, such as flow analysis, queuing theory and simulation rely on process execution data. Furthermore process execution data is needed to perform process mining. Process mining refers to the

extraction of knowledge on event logs to discover, monitor and improve processes[Vai13] which can also be performed on data from systems which do not execute a explicit process model, e.g. ticketing or Enterprise Resource Planning (ERP) systems.

The development of BPM went hand in hand with technological progress. “Information technology in general and information systems in particular deserve an important role in business process management, because more and more activities that a company performs are supported by information systems” [Wes12]. Different types of IT system emerged, most notably Enterprise Resource Planning (ERP) systems and Workflow Management Systems (WfMSs), which enabled a central data retention and easy control but also monitoring of business processes [DRMR13]. Already in 1990, Davenport highlighted the importance of IT for process-centered management and described a “recursive relationship between IT capabilities and business process redesign” [DES89].

For WfMSs relevant process execution data[1] is usually stored in Database Management Systems (DBMSs). Workflow control data in particular, but also so called worfkflow relevant data, is “necessary for the operation of the workflow and the realisation of routing“[Jó06]. Whereas workflow relevant data may be manipulated by workflow applications as well as by the WfMS, workflow control data is exclusively ”managed by the WfMS and/or a workflow engine. Such data is internal to the WfMS and is not normally accessible to applications“[Wor99].

Temporal aspects in data are crucial for workflow control data and workflow relevant data, but also for event logs of process executions. However, even though there was a growing need to handle temporal data already in the mid-90s when BPM arose, there was only a weak support of temporal aspects as part of database standard for quite a long time[Zan08]. E.g. the standardization of temporal data types, temporal queries and support to store meta data such as valid time, transaction time or decision time was introduced to ISO standard for Structured Query Language (SQL) in 2011[ISO11].
Even though standards have been extended and some conventional ”SQL-based DBMSs are capable of supporting the management of interval data, the support they offer can be improved considerably“. This was proved by the Free University of Bozen-Bolzano which adapted the evaluation engine of a common DBMS. The approach was to reduce temporal queries to non-temporal queries over data with adjusted interval [DBGJ16].

[1] [Wor99] defines application data as one kind of BPM relevant data. Application data is only known to evoked applications and not relevant for process control and therefore not in focus of the thesis.

CHAPTER 5

Relevancy to the curriculum of Business Informatics

The topic of Business Process Management (BPM) is part of the master programme Business Informatics. Thus, the goal of acquiring the according skills and abilities during the studies is documented in the curriculum, e.g. “Graduates are among others qualified for the following fields of activity: Analysis and optimization [...] of business processes and according information processes“. Furthermore, this thesis compares two different database technologies, which is also part of the curriculum, e.g. ”decision making between competitive architectures and designs” [Cur11].

The relevancy of business processes, BPM and Database Management System technologies within the curriculum of Business Informatics is also evidenced by their occurrence in various lectures, such as:

- 188.924 Workflow Modeling and Process Management
- 188.405 Advanced Software Engineering
- 188.483 Knowledge Management
- 330.232 IT-based Management
- 188.427 E-Commerce

Acronyms

BPM Business Process Management. 4, 6–8

BPMN Business Process Model and Notation. 6

DBMS Database Management System. 7, 8

ERP Enterprise Resource Planning. 7

IASB International Accounting Standards Board. 1

IFRS International Financial Reporting Standard. 1

SQL Structured Query Language. 7

WfMS Workflow Management System. 1, 7

Bibliography

[Cur11] Studienplan 2011 für das Masterstudium Business Informatics an der Technischen Universität Wien. http://www.informatik.tuwien.ac.at/studium/angebot/studienplaene/066926-SP-2011-Mas_Business_Informatics.pdf, October 2011.

[DBGJ16] Anton Dignös, Michael H. Böhlen, Johann Gamper, and Christian S. Jensen. Extending the kernel of a relational dbms with comprehensive support for sequenced temporal queries. *ACM Trans. Database Syst.*, 41(4):26:1–26:46, November 2016.

[DES89] Thomas Davenport and James E. Short. The new industrial engineering: Information technology and business process redesign. *Sloan management Review*, 31, November 1989.

[DRMR13] Marlon Dumas, Marcello La Rosa, Jan Mendling, and Hajo A. Reijers. *Fundamentals of Business Process Management, First Edition.* Springer, 2013.

[Int14] International Accounting Standards Board. IFRS 9 Financial Instruments (replacement of IAS 39). https://www.ifrs.org/issued-standards/list-of-standards/ifrs-9-financial-instruments/, July 2014. Draft.

[ISO11] ISO. *ISO/IEC 9075-1:2011 Information technology — Database languages — SQL.* December 2011.

[Jó06] Tick József. Workflow model representation concepts. In *International Symposium of Hungarian Researchers on Computational Intelligence*, number 7, pages 329–337, January 2006.

[Sno00] R.T. Snodgrass. *Developing Time-oriented Database Applications in SQL.* Data Management Systems Series. Morgan Kaufmann Publishers, 2000.

[Vai13] A. Vaisman. An introduction to business process modeling. *Lecture Notes in Business Information Processing*, 138:29–61, 2013.

[Wes12] Mathias Weske. *Business Process Management: Concepts, Languages, Architectures*. Springer, Berlin, 2 edition, 2012.

[Wor99] Workflow Management Coalition. Terminology and glossary. Document Number WFMC-TC-1011 3.0. `http://www.wfmc.org/docs/TC-1011_term_glossary_v3.pdf`, February 1999.

[Zan08] Carlo Zaniolo. Time versus standards: A tale of temporal databases. In Il-Yeol Song, Mario Piattini, Yi-Ping Phoebe Chen, Sven Hartmann, Fabio Grandi, Juan Trujillo, Andreas L. Opdahl, Fernando Ferri, Patrizia Grifoni, Maria Chiara Caschera, Colette Rolland, Carson Woo, Camille Salinesi, Esteban Zimányi, Christophe Claramunt, Flavius Frasincar, Geert-Jan Houben, and Philippe Thiran, editors, *Advances in Conceptual Modeling – Challenges and Opportunities*, pages 67–67, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.