

Seyed-Mehdi-Reza Beheshti · Boualem Benatallah
Sherif Sakr · Daniela Grigori
Hamid Reza Motahari-Nezhad · Moshe Chai Barukh
Ahmed Gater · Seung Hwan Ryu

Process Analytics

Concepts and Techniques for Querying
and Analyzing Process Data



Springer

Process Analytics

Seyed-Mehdi-Reza Beheshti • Boualem Benatallah •
Sherif Sakr • Daniela Grigori •
Hamid Reza Motahari-Nezhad • Moshe Chai Barukh •
Ahmed Gater • Seung Hwan Ryu

Process Analytics

Concepts and Techniques for Querying
and Analyzing Process Data

Seyed-Mehdi-Reza Beheshti
The University of New South Wales
Sydney, NSW, Australia

Boualem Benatallah
The University of New South Wales
Sydney, NSW, Australia

Sherif Sakr
The University of New South Wales
Sydney, NSW, Australia

Daniela Grigori
Laboratoire LAMSADE
Université Paris Dauphine
Paris, France

King Saud bin Abdulaziz University
for Health Sciences
Riyadh, Saudi Arabia

Hamid Reza Motahari-Nezhad
IBM Almaden Research Center
San José, CA, USA

Moshe Chai Barukh
The University of New South Wales
Sydney, NSW, Australia

Ahmed Gater
Ikayros
Paris, France

Seung Hwan Ryu
The University of New South Wales
Sydney, NSW, Australia

ISBN 978-3-319-25036-6

ISBN 978-3-319-25037-3 (eBook)

DOI 10.1007/978-3-319-25037-3

Library of Congress Control Number: 2015956171

© Springer International Publishing Switzerland 2016

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Foreword

The quality of the services any organization provides largely depends on the quality of their processes. Companies are increasingly grasping this concept and moving towards process-oriented enterprises. As this happens, the central problem (and opportunity) for any organization becomes the identification, measurement, analysis, and improvement of its processes. Without this, it is very difficult—if not impossible—to be competitive.

This book provides researchers and practitioners with invaluable knowledge in the area of process management in general and process analysis in particular. The book starts from an introduction on process modeling and process paradigms, which is very easy to understand for novices but that also includes interesting bits of information for people who are experts in the area. From there, it proceeds to teaching how you can query and analyze process models and moves on to the analysis of process execution data. In this way, you get a complete picture of what you need to do to identify, understand, and improve business processes.

The book is very well written and a pleasure to read. It presents concepts that are normally not easy to grasp in a manner that is simple and intuitive and also a way to classify and structure topics that makes it easy to make sense out of a rather large set of concepts, paradigms, and techniques. All authors have many years of experience in Business Process Management (BPM) and, thanks to this and to a combined academia-industry background, they have been clearly able to identify which are the key, hard analytics problems in BPM that are relevant in practice and they guide the readers to show them how these problems can be solved concretely.

What I particularly enjoyed about reading this book, and I am sure you will enjoy as well, is that it deals with very realistic environments, where analysis is complex: in any company, you have very different systems and very different data, all related to the same processes. Indeed, most organizations have accumulated systems that support different subprocesses over the years. This means that if you really want to understand what is going on in your business and how to make it better, you need to get to the individual data source, correlate the data, follow the breadcrumbs, and then work up to discover the process model and its key performance indicators. Performing all of this, potentially in an explorative manner, needs process data

querying, data correlation, data analysis, as well as process model-level analysis and matching techniques described in this book. Through such a journey, you will have a clearer picture of how much each process takes in terms of time, resources, and cost and where the “bottlenecks” are with regard to efficiency and effectiveness, which will help understand where and how to improve. This is an aspect that many authors forget or neglect, but with this book you will learn relatively simple but effective ways of approaching such scenarios.

I hope you like the book as much as I did. Enjoy your reading!

Trento, Italy
2015

Fabio Casati

Preface

Business processes are inseparable from data—data from the execution of business processes, documentation and description of processes, process models, process variants, artifacts related to business processes, and data generated or exchanged during process execution. Process data can be in various forms from structured to semistructured and unstructured. A variety of data capturing, collection, and instrumentation tools and process implementation over various types of systems have amplified the amount of data that are gathered and are available about processes. Business process monitoring, analysis, and intelligence techniques have been investigated in depth over the last decade to enable delivering process insight and support analytic-driven decision making about processes. However, they have been developed under the assumption of the existence of a workflow engine as the central repository for process data and for process analysis. As processes touch almost every system and leave an operational data footprint everywhere in an enterprise, and the wide span of process support the majority of information and enterprise application systems, the existing tools and methods have proved to be inadequate and limiting for correlating, managing, and analyzing process data.

Considering various process analytics needs and the importance of process analytics over a myriad of process data forms, formats, and levels of abstractions, we felt the need for writing this book as a technical introduction to the field of process analytics to share the state of the art in research achievements and practical techniques of process analytics. Given that many “data analysis” subjects such as querying, matching, and warehousing are covered well in the literature, the focus of this book will be on concepts, techniques, and methods for analyzing process data and offering a common understanding of the concepts and technologies in process analytics. This book will cover a large body of knowledge in process analytics, including process data querying, analysis, and matching and correlating process data and models to assist practitioners as well as researchers in understanding underlying concepts, problems, methods, tools, and techniques for modern process analytics. It starts by introducing basic business process and process analytics concepts, describes the state of the art in this space, and continues by taking a deeper perspective on various analytics techniques covering analytics over different levels

of process abstractions, from process execution data to methods for linking and correlating process execution data and to inferring process models and querying process execution data and process models and scalable process data analytics methods. The book also provides a review of commercial process analytics tools and their practical applications.

Chapters Overview

In particular, we introduce concise and commonly accepted definitions, taxonomies, and frameworks for business process management in Chap. 2. We briefly discuss process-centered systems, process modeling, and why process model developers have to model collaboratively in today's world. We classify strategies that a business may use to automate processes. We discuss technologies, applications, and practices used to provide business analytics. We discuss a wide spectrum of business process paradigms that has been presented in the literature from structured to unstructured processes. We discuss state-of-the-art technologies and demystify the concepts, abstractions, and methods in structured and unstructured BPM, including activity-, rule-, artifact-, and case-based processes.

Chapter 3 presents process matching techniques that are useful in a series of process model analytics tasks. We present a brief overview of schema matching techniques as they are useful for organizing process data, more precisely for the integration of process execution data from various, potentially heterogeneous, systems and services in a process event log. Besides integrating process execution data, process models from various repositories (or resulting from process mining tasks) have to be assembled in the process model layer of the process space (see Fig. 1.3). The management of such integrated repositories requires techniques to manage versions (operators like merge, diff) and for duplicate detections that are based on process matching. We also classify approaches for matching different dimensions of process models (data, interface, protocol, etc.) and analyze their strengths and limitations. We also discuss that process matching techniques are needed for conformance evaluation (comparing the process model mined from event logs and the initially designed process model or its description in textual documentation). Process similarity measures presented in Chap. 3 are also used for clustering process models and for organizing process models mined from event logs for unstructured processes, like those in the medical domain.

Querying is an important aspect in business process analytics. In particular, querying techniques are commonly used in the different phases of business process life cycles for various analytics purposes. For example, during the design time, querying techniques can be used to retrieve existing process models, check the similarity between new process models and existing processes, and conduct compliance checks against existing business logic or rules. During the runtime phase of the business process life cycles, querying techniques can be utilized for analyzing and understanding the execution patterns of business processes. In addition, they can

be used to instantly detect any compliance violations. In offline mode, querying techniques can be used for auditing, mining, and process improvement purposes. To address the importance of querying business processes, in Chap. 4 we focus on business process querying techniques and languages. We describe the key concepts, methods, and languages for querying business processes. We describe the foundation and enabling technology for understanding the execution of a business process in terms of its scope and details, which is a challenging task and requires querying process repositories and business process execution.

Identifying business needs and determining solutions to business problems require the analysis of business process data. In Chap. 5, we discuss that the analysis of business data will help in discovering useful information, suggesting conclusions, and supporting decision making for enterprises. In Chap. 5, we give an overview of the different aspects of business data analysis techniques and approaches from process/data spaces to data provenance and data-based querying techniques. We provide an overview of *warehousing process data* followed by an introduction to *data services* and *DataSpaces*, which facilitate organizing and analyzing process-related data. We discuss the importance of supporting big data analytics over process execution data. We then take a holistic view of the process executions over various information systems and services (i.e., process space) followed by a brief overview of process mining to highlight the interpretation of the information in the enterprise in the context of process mining. Moreover, we focus on introducing cross-cutting aspects in process data and discuss how process analytics can benefit from cross-cutting aspects such as provenance, e.g., to analyze the evolution of business artifacts.

Finally, Chap. 6 provides an overview of open-source and commercial software for process analytics. Software for process analytics can be applied to the rich source of events that document the execution of processes and activities within BPM systems in order to support decision-making in organizations. In this context, various existing tools focus on the behavior of completed processes, evaluate currently running process instances, or focus on predicting the behavior of process instances in the future. In this chapter, we provide a summary and comparison of existing open-source and commercial software for process analytics, including real-world use case scenarios, followed by a discussion and future directions on some of the emerging and hot trends in the business process management area such as process spaces, big data for processes, crowdsourcing, social BPM, and process management on the cloud. We briefly describe the core essence of these directions and discuss how they can facilitate the analysis of business processes.

Who Is This Book For?

In writing this book, we have taken into consideration a wide range of audience interested in the space of business process management and process analytics. We have tried to cover topics of interest for academics (professors, researchers, and

research students), professionals (managers, data scientists, analysts, and software engineers), and practitioners in understanding and employing process analytics methods and tools to gain a better insight within and across business processes. This book is a comprehensive textbook on process analytics and, therefore, could be a useful reference for academics, professionals, and practitioners.

To Professors You will find this book useful for a variety of courses, from an undergraduate course in business process management up through a graduate course in business process analytics. We have provided considerably more material than can fit in a typical one-term course; therefore, you can think of the book as a buffet from which you can pick and choose the material that best supports the course you wish to teach.

To Research Students and Researchers We hope that this textbook provides you with an enjoyable introduction to the field of process analytics and business process management. We have attempted to properly classify the state of the art, describe technical problems and techniques/methods in depth, and highlight future research directions in process analytics.

To Professionals and Practitioners You will find this book useful as it provides a review of the state of the art in commercial tools and techniques and also describes real-world use case scenarios. The wide range of topics covered in this book makes it an excellent handbook on process analytics. Most of the methods that we discuss in each chapter have great practical utility: process matching techniques in Chap. 3, process querying techniques and languages in Chap. 4, and process data analysis in Chap. 5. Therefore, in each chapter you can find details for the state-of-the-art tools and methods. An overview of open-source and commercial software for process analytics is finally provided in Chap. 6.

Sydney, New South Wales, Australia
 Sydney, New South Wales, Australia
 Sydney, New South Wales, Australia
 Paris, France
 San José, California, USA
 Sydney, New South Wales, Australia
 Paris, France
 Sydney, New South Wales, Australia

Seyed-Mehdi-Reza Beheshti
 Boualem Benatallah
 Sherif Sakr
 Daniela Grigori
 Hamid Reza Motahari-Nezhad
 Moshe Chai Barukh
 Ahmed Gater
 Seung Hwan Ryu

Contents

1	Introduction	1
1.1	The Modern Enterprises and the Need for Process Analytics	1
1.2	Business Processes: An Overview	3
1.2.1	Business Processes in Modern Enterprises	7
1.3	Process Analytics	11
1.3.1	Capturing Process Data	11
1.3.2	Organizing Process Data	13
1.3.3	Process Space	14
1.3.4	Business Process Analytics	15
1.4	Goals, Structure, and Organization	17
2	Business Process Paradigms	19
2.1	Introduction	19
2.2	Dimensions for Characterizing Processes	21
2.2.1	Process Paradigms	21
2.2.2	Process Representation Models/Languages	23
2.2.3	Process Implementation Technologies	23
2.3	Process Implementation Technologies	24
2.3.1	Workflow Engines	24
2.3.2	Rule Engines	25
2.3.3	Program Coded	31
2.4	Process Paradigms: A Survey of Frameworks and Tools	33
2.4.1	Structured Processes	33
2.4.2	Unstructured (Ad Hoc) Processes	48
2.4.3	Case-Management (Semi-structured) Processes	56
2.5	Conclusions and Future Directions	59
3	Process Matching Techniques	61
3.1	Introduction	61
3.1.1	Application Domains	62
3.1.2	Similarity Metrics Used When Comparing Process Models	64

3.2	Schema Matching	66
3.2.1	Individual Matcher Approaches	67
3.2.2	Multi-matcher Matching Approaches	67
3.3	Interface Matching	68
3.3.1	Matching Operations	69
3.3.2	Matching QoS Factors	71
3.4	Protocol Matching	73
3.5	Process Matching	76
3.5.1	Abstract Models for Process Model Descriptions	78
3.5.2	Matching Process Activities	80
3.5.3	Structural Matching	82
3.5.4	Behavioral Matching	85
3.6	Discussion and Concluding Remarks	87
4	Model-Based Business Process Query Techniques and Languages	91
4.1	Querying Repositories of Business Process Models	92
4.1.1	BPMN-Q	93
4.1.2	BP-QL	96
4.1.3	BeehiveZ	97
4.1.4	APQL	98
4.2	Querying Business Process Execution	99
4.2.1	BP-Mon	99
4.2.2	BP-SPARQL	101
4.3	Business Process Compliance	101
5	Business Process Data Analysis	107
5.1	Warehousing Business Process Data	108
5.1.1	Data Services	109
5.1.2	DataSpaces	110
5.2	Supporting Big Data Analytics over Process Execution Data	113
5.2.1	Online Analytical Processing	115
5.2.2	Trend, What-If, and Advanced Analysis	116
5.3	Business Data Analytics and Process Spaces	118
5.3.1	Process Space	119
5.3.2	Logical Components of Process Spaces	121
5.3.3	Process Space Management System	123
5.4	Process Mining	127
5.5	Analyzing Crosscutting Aspects in Processes' Data	131
5.6	Provenance and Evolution of Business Artifacts	132
6	Tools, Use Cases, and Discussions	135
6.1	Observations	135
6.2	Open-Source and Commercial Tools for Process Analytics	137
6.2.1	BPM in the Cloud	138
6.2.2	Business Process Analytics: Practical Use Case Scenario	142

- 6.3 Discussion and Future Directions 144
 - 6.3.1 Big Data Analytics for Process Data 144
 - 6.3.2 Analyzing Big Process Data Problem 146
 - 6.3.3 Crowdsourcing and Social BPM 148
 - 6.3.4 Process Data Management in the Cloud 149
- References** 151
- Index** 177

List of Figures

Fig. 1.1	An example of a business process execution in modern enterprises	2
Fig. 1.2	An example of the ad-hoc business process execution in an enterprise	8
Fig. 1.3	Process-related abstraction layers and process analytics: how analytics can help in understanding business processes in modern enterprises	15
Fig. 1.4	Throughout the different chapters of this book, we will dive into detail with respect to the different approaches for analyzing business process	18
Fig. 2.1	Dimensions for characterizing processes	22
Fig. 2.2	Workflow engine execution procedure (typically: scheduling and resource assignment)	25
Fig. 2.3	Production rule-engine typical architecture and execution procedure	29
Fig. 2.4	Artifact-centric conceptual and technological landscape.....	37
Fig. 2.5	(a) Requirement life cycle for software development artifacts; (b) RACI matrix for specification of roles and Responsibilities	39
Fig. 2.6	Event-Driven Business Process Management high-level architecture	46
Fig. 2.7	Integration of rule engine with BPM systems: (a) stateless versus (b) stateful (knowledge-session) approach	47
Fig. 2.8	Unstructured process-support landscape	49
Fig. 3.1	Example of annotated FSM intersection-based matching	75
Fig. 3.2	Process model mismatches	77
Fig. 3.3	Structural vs. behavioral matching	78
Fig. 3.4	Example of a process specification using (a) FSM, (b) PG, and (c) WF-Nets.....	80

Fig. 3.5 Granularity-level differences 81

Fig. 4.1 Basic BPMN-Q constructs 93

Fig. 4.2 An example BPMN-Q query with a match to a process model 94

Fig. 4.3 Query to look for loan handling processes 95

Fig. 4.4 Queries semantically similar to the one in Fig. 4.3 95

Fig. 4.5 Framework architecture for business process compliance management 102

Fig. 4.6 Business process compliance monitoring with SARI 105

Fig. 4.7 The architecture of the SeaFlows toolset 106

Fig. 5.1 An example of the business process execution in modern enterprises 119

Fig. 5.2 Process views defined over information items characterize a process space 122

Fig. 5.3 The components of a process space management system 123

Fig. 5.4 Positioning of the three main types of process mining: (a) discovery, (b) conformance checking, and (c) enhancement 129

Fig. 5.5 Example case scenario for breast cancer treatment including a case instance (a); parent artifacts, i.e., ancestors, for a patient history document (b) and its versions (c); and a set of activities which shows how version v_2 of a patient history document develops and changes gradually over time and evolves into version v_3 (d) 133

Fig. 6.1 Research directions in understanding big process data 145

Chapter 1

Introduction

Business processes are central to the operation of public and private enterprises. For most enterprises, the success is strictly related to how efficient and effective the execution of their processes is. For this reason, business process analytics has always been a key endeavor for companies. Early efforts to address this goal started with process *automation* where workflow and other middleware technologies were used to reduce human involvement by better systems integration and automated execution of business logic. The total or partial automation of the process creates an unprecedented opportunity to gain visibility on process executions in the enterprise. Recently, the focus of process thinking has shifted toward understanding and analyzing business processes and business process-related data captured in various information systems and services that support processes. In this chapter, we give an overview of the different approaches for analyzing business processes at the different phases of its life cycle.

1.1 The Modern Enterprises and the Need for Process Analytics

Information processing using knowledge-, service-, and cloud-based systems has become the foundation of twenty-first-century life. These systems run and support processes in our governments, industries, transportations, and hospitals and even our social life. In this context, business processes (BPs) and their continuous improvements are key to the operation of any system supporting our life and in enterprises. Typical examples of BPs that are supported by systems include those that automate the operation of commercial enterprises such as banking and financial transaction processing systems.

Over the last decade, many BPs across and beyond the enterprise boundaries have been integrated. Process data is stored across different systems, applications, and services in the enterprise and sometimes shared between different enterprises

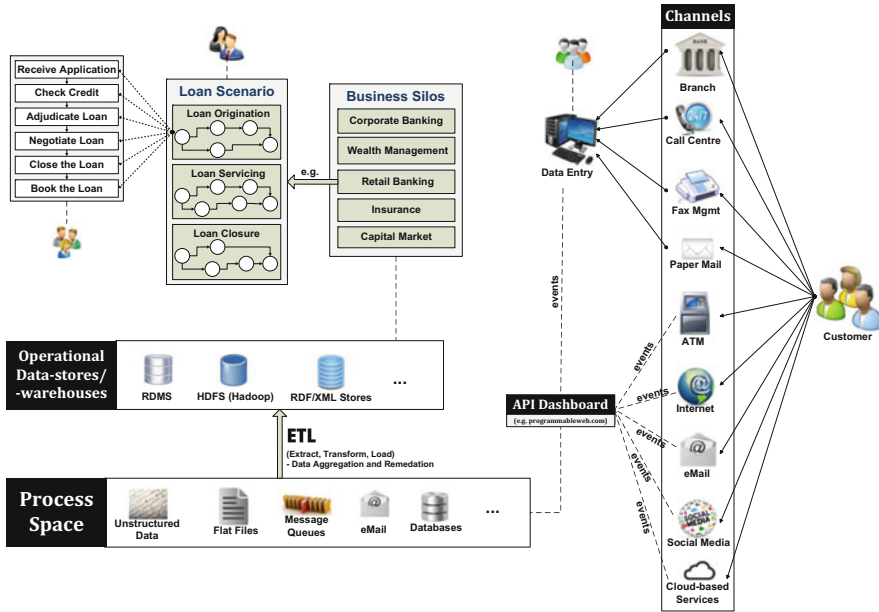


Fig. 1.1 An example of a business process execution in modern enterprises

to provide the foundation for business collaborations. Figure 1.1 illustrates a business process scenario in the banking context that spans across multiple systems and services inside and among third-party services providers. These systems are distributed over various networks, but when viewed at a macrolevel, various organizations and systems are components of a larger, logically coherent system. Let us have a closer look at this example scenario.

Figure 1.1 demonstrates a property lending scenario in the banking system. Consider Adam (a customer) who plans to buy a property. He needs to find a lending bank. He can use various crowdsourcing services (e.g., Amazon Mechanical Turk¹) or visit a mortgage bank to find candidate banks. Then he needs to contact the banks through one of many channels and start the loan pre-approval process. After that, he may visit various Web sites or real-estate services to find a property. Meanwhile, he can use social network Web sites (e.g., Facebook or Twitter) to socialize the problem of buying a property and ask for others' opinions on finding a suburb. After finding a property, he needs to get a solicitor to start the process of buying the property. Lots of other processes can be executed before or after this point. For example, the bank may outsource the process of confirming Adam's income, during pre-approval process, to other companies or appraising the property.

¹<https://www.mturk.com/>

As can be seen, the data relevant to the business process of a bank is scattered across multiple systems, and in many situations, stakeholders can be aware of processes, but they are not able to track or understand it: it is important to maintain the vital signs of bank processes by analyzing the process data. This task is challenging as a huge amount of data needs to be processed (it requires scalable methods) and enables answering process-centric queries, e.g., “Where is loan-order #756? What happened to it? Is it blocked? Give me all the documents and information related to the processing of loan-order #756. What is the typical path of loan orders? What is the process flow for it? What are the dependencies between loan applications A_1 and A_2 ? How much time and resources are spent in processing loan orders that are eventually rejected? Can I replace X (i.e., a service or a person) by Y? Where did the data come from? How was it generated? Who was involved in processing file X? At which stage do loan orders get rejected? How many loan orders are rejected between the time period τ_1 and τ_2 ? What is the average time spent on processing loan orders? Where do delays occur?” Can the bank outsource the process of confirming Adam’s income or appraising the property to provider X (i.e., is the business protocol of the provider X compatible with the process model of the bank)? What are the differences between public process models of two different banks?

Under such conditions, *process analytics* becomes of a great practical value but clearly a very challenging task as well.

1.2 Business Processes: An Overview

A business process (BP) is a set of coordinated tasks and activities, carried out manually or automatically, to achieve a business objective or goal [6, 346]. An activity is the smallest unit of work performed by executing a program, enacting a human or machine action, or invoking another business process (known as subprocess) [7, 346]. A business process is typically *structured* and, therefore, is associated with a data flow, showing what data and how they are transferred among activities and tasks, and a control flow, showing the order in which activities and tasks are performed.

Two types of business processes are recognized: public and private. Public business processes can be shared with business partners (e.g., clients and suppliers) within an enterprise and can be used in the business-to-business integration (B2Bi) context [29]. On the contrary, private business processes are internal to the enterprise, include execution details, and can be used in enterprise application integration (EAI) context [29]. In order to manage organization performance through BPs, a set of methods, techniques, and tools are defined, developed, and presented under a concept known as business process management (BPM).

A BPM system is a “generic software system that is driven by explicit process designs to enact and manage operational business processes” [7], where operational processes refer to repetitive business processes performed by organizations in