Data Warehouse Requirements Engineering

Naveen Prakash · Deepika Prakash

Data Warehouse Requirements Engineering

A Decision Based Approach

Springer

Naveen Prakash
ICLC Ltd.
New Delhi
India

Deepika Prakash
Central University of Rajasthan
Kishangarh
India

ISBN 978-981-10-7018-1 ISBN 978-981-10-7019-8 (eBook)
https://doi.org/10.1007/978-981-10-7019-8

Library of Congress Control Number: 2017961755

Printed on acid-free paper

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd. part of Springer Nature.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

To
Our Family

Preface

That requirements engineering is part of the systems development life cycle and is about the first activity to be carried out when building systems is today considered as basic knowledge in computer science/information technology. Requirements engineering produces requirements specifications that are carried through to system design and implementation. It is assumed that systems automate specific activities that are carried out in the real world. These activities are transactions, for example reservations, cancellations, buying, selling, and the like. Thus requirements engineering produces requirements specifications of transactional systems.

So long as systems were not very complex, the preparation of a requirements specification was feasible and did not compromise on system delivery times. However, as systems became more and more complex, iterative and incremental development came to the fore. Producing a requirements specification is now frowned upon and we need to produce, in the language of Scrum, user stories for small parts of the system.

About the time requirements engineering was developing, data warehousing also became important. Data warehouse development faced the same challenges as transactional systems do, namely determination of the requirements to be met and the role of requirements engineering in the era of agile development. However, both these issues have been taken up relatively recently.

Due to this recent interest in the area, requirements engineering for data warehousing is relatively unknown. We fear that there is widespread paucity of understanding of the nature of data warehouse requirements engineering, how it differs from traditional transaction-oriented requirements engineering and what are the new issues that it raises.

Perhaps, the role of agility in data warehouse development is even more crucial than in transactional systems development. This is because of the inherent complexity of data warehouse systems, long lead times to delivery, and the huge costs involved in their development. Indeed, the notion of data marts and the bus approach to data warehouse development is an early response to these challenges.

This book is our attempt at providing exposure to the problem of data warehouse requirements engineering. We hope that the book shall contribute to a wider

awareness of the difference between requirements engineering for transactional and data warehouse systems, and of the challenges that data warehousing presents to the requirements engineering community.

The position adopted in this book is that even in the face of agile development, requirements engineering continues to be relevant. Requirements engineering today is not to produce requirements specifications of entire systems. Rather, it is done to support incremental and iterative development. In other words, rather than restrict incremental and iterative development to downstream tasks of design and implementation, we must extend it to the requirements engineering task as well. We argue that the entire data warehouse systems development life cycle should become agile.

Thus, we make requirements and requirements engineering as the fulcrum for data warehouse agile development. Just as requirements specifications *of systems* formed the basis for proceeding with systems development earlier, so also now requirements specifications *of system increments* must form the basis of incremental and iterative development.

Following this line of argument, instead of a requirements specification, we propose to develop requirements granules. It is possible to consider building a requirements granule per data mart. However, we consider a data mart as having very large granularity because it addresses an entire subject like sales, purchase etc. Therefore, the requirements granule that will be produced shall be large-grained resulting in relatively long lead times to delivery of the intended product increment. It is worth developing an approach to requirements engineering that can produce requirements granules of smaller sizes.

To reduce the sizes of requirements granules, we introduce the notion of a decision and propose to build data warehouse fragments for decisions. Thus, data warehouse requirements engineering is for discovering the decisions of interest and then determining the information relevant to each decision. A requirements granule is the collection of information relevant to a decision. If this information is available then it is possible for the decision maker to obtain it from the data warehouse fragment, evaluate it, and decide whether to take the decision or not. This implies that the size of a granule is determined by the amount of information that is associated with a decision.

The notion of a decision is thus central to our approach. A decision represents the useful work that the data warehouse fragment supports and a data warehouse fragment is the implementation of a requirements fragment. The approach in this book represents a departure from the conventional notion of a data mart that is built to "analyze" a subject area. Analysis for us is not an aim in itself but taking a decision is and analysis is only in support of the decision making task.

As more and more decisions are taken up for development, there is a proliferation of requirements granules and data warehouse fragments. This results in problems of inconsistent information across the enterprise, and of proliferating costs due to multiple platforms and ETL processes. This is similar to what happens in the bus-of-data-marts approach except that a decision may be of a lower granularity than a data mart. This means that we can expect many more data warehouse

fragments than data marts and the problem of inconsistency and costs is even more severe.

Given the severity of the problem, we do not consider it advisable to wait for the problem to appear and then take corrective action by doing consolidation. It is best to take a preventive approach that minimizes fragment proliferation. Again, keeping in mind that for us requirements are the fulcrum for data warehouse development, we consolidate requirements granules even as they are defined.

This book is a summary of research in the area of data warehouse requirements engineering carried out by the authors. To be sure, this research is ongoing and we expect to produce some more interesting results in the future. However, we believe that we have reached a point where the results we have achieved form a coherent whole from which the research and industrial community can benefit.

The initial three chapters of the book form the backdrop for the last three. We devote Chap. 1 to the state of the art in transactional requirement engineering whereas Chap. 2 is for data warehouse requirements engineering. The salient issues in data warehouse requirements engineering addressed in this book are presented in Chap. 3.

Chapter 4 deals with the different types of decisions and contains techniques for their elicitation. Chapter 5 is devoted to information elicitation for decisions and the basic notion of a requirements granule is formulated here. Chapter 6 deals with agility built around the idea of the requirements granules and data warehouse fragments. The approach to data warehouse consolidation is explained here.

The book can be used in two ways. For those readers interested in a broad-brush understanding of the differences between transactional and data warehouse requirements engineering, the first three chapters would suffice. However, for those interested in deeper knowledge, the rest of the chapters would be of relevance as well.

New Delhi, India — Naveen Prakash
Kishangarh, India — Deepika Prakash
September 2017

Contents

About the Authors

Naveen Prakash started his career with the Computer Group of Bhabha Atomic Research Centre Mumbai in 1972. He obtained his doctoral degree from the Indian Institute of Technology Delhi (IIT Delhi) in 1980. He subsequently worked at the National Center for Software Development and Computing Techniques, Tata Institute of Fundamental Research (NCSDCT, TIFR) before joining the R&D group of CMC Ltd where he worked for over 10 years doing industrial R&D. In 1989, he moved to academics. He worked at the Department of Computer Science and Engineering, Indian Institute of Technology Kanpur (IIT Kanpur), and at the Delhi Institute of Technology (DIT) (now Netaji Subhas Institute of Technology (NSIT)), Delhi. During this period he provided consultancy services to Asian Development Bank and African Development Bank projects in Sri Lanka and Tanzania, respectively, as well as to the Indira Gandhi National Centre for the Arts (IGNCA) as a United Nations Development Programme (UNDP) consultant. He served as a scientific advisor to the British Council Division, New Delhi and took up the directorship of various educational institutes in India. Post-retirement, he worked on a World Bank project in Malawi.

Prof. Prakash has lectured extensively in various universities abroad. He is on the editorial board of the Requirements Engineering Journal, and of the International Journal of Information System Modeling and Design (IJISMD). He has published over 70 research papers and authored two books.

Prof. Prakash continues to be an active researcher. Besides Business Intelligence and Data Warehousing, his interests include the Internet-of-things and NoSQL database. He also lectures at the Indira Gandhi Delhi Technical University for Women (IGDTUW), Delhi and IIIT Delhi.

Deepika Prakash obtained her Ph.D. from Delhi Technological University, Delhi in the area of Data Warehouse Requirements Engineering. Currently, she is an Assistant Professor at the Department of Big Data Analytics, Central University of Rajasthan, Rajasthan.

Dr. Prakash has five years of teaching experience, as well as two years of experience in industrial R&D, building data marts for purchase, sales and inventory and in data mart integration. Her responsibilities in industry spanned the complete life cycle, from requirements engineering through conceptual modeling to extract-transform-load (ETL) activities.

As a researcher, she has authored a number of papers in international forums and has delivered invited lectures at a number of Institutes throughout India. Her current research interests include Business Intelligence, Health Analytics, and the Internet-of-Things.