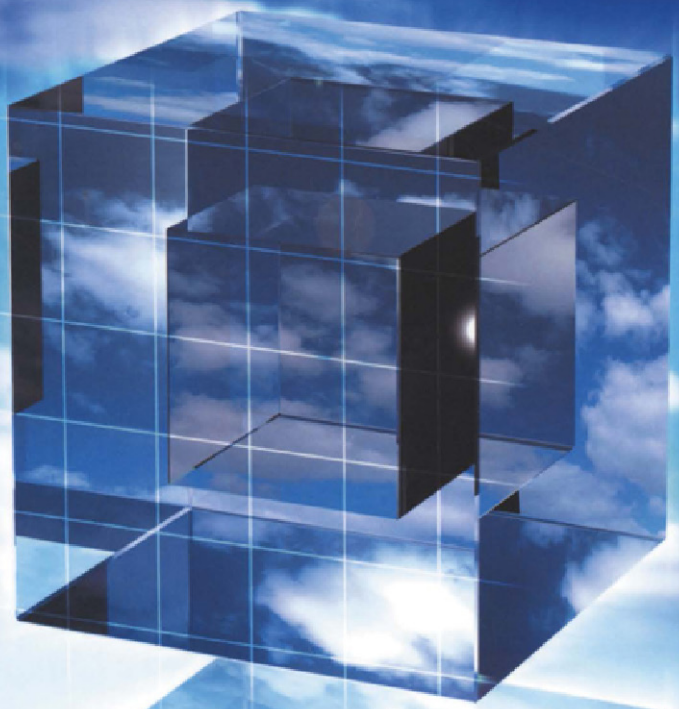
For SQL Server 2000

Fast Track to MDX

Mark Whitehorn
Robert Zare • Mosha Pasumansky

Second Edition

Introduces Recursion



EXTRAS ONLINE

Springer

Fast Track to MDX

Fast Track to MDX

Second edition

Mark Whitehorn, Robert Zare
and Mosha Pasumansky

Springer

Mark Whitehorn
University College Worcester, Worcester, UK

Robert Zare
Microsoft Corporation, Redmond, WA, USA

Mosha Pasumansky
Microsoft Corporation, Redmond, WA, USA

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Control Number: 2005933165

Additional material to this book can be downloaded from http://extra.springer.com.

ISBN-10: 1-84628-174-1 Printed on acid-free paper
ISBN-13: 978-1-84628-174-7

ISBN 1-85233-681-1 1st edition

First published 2002
Second edition 2006

Printed in the United States of America (IK/MV)

9 8 7 6 5 4 3 2 1

Springer Science+Business Media
springeronline.com

Contents

Foreword

I love Business Intelligence. I love BI because it is all about becoming better. BI is all about empowering us with knowledge and that knowledge is the power to realize our full potential. As Zorge the spy said, "knowledge is power", and who doesn't love to have the power to know, to understand and to make intelligent decision? I do.

Since the dawn of the modern information system it was obvious that the information accumulated in the machine is wasted if there is no way to analyze it and learn from it. From as early as the 1950s, data analysis systems and, later, decision support systems were designed, developed and deployed with that intent. However, only in the last decade have these systems become both reasonably affordable and mainstream and their business impact indisputable.

The last decade has also seen the emergence of OLAP as the centerpiece of the BI technologies. The OLAP multidimensional databases combine incredible performance with unsurpassed analytical power and, in my opinion, are the foundation of the BI platform.

While the performance differences between the multidimensional databases and the traditional relational databases are very significant, Moore's law, which states that the hardware computing power doubles every 18 months, renders this advantage of the OLAP databases temporary. Sooner or later, the raw computing power of the common server machines will be sufficient to provide the performance needed for sophisticated analysis even when the data is stored in a relational database.

However, where OLAP is likely to maintain a sustained advantage over the relational database is in its analytical capabilities. Here the differences are much deeper. The multidimensional data model is vastly superior to the relational data model when it comes to the expressiveness of analytical operations. The ability to have random access to any point in space, both detailed data and aggregates, makes it a breeze to express calculations that would otherwise take pages of SQL statements using a relational database.

This is also where MDX enters the picture and why MDX is so important. MDX is the way to express the analytics in the OLAP database. Without

MDX, all of the sophisticated calculations and smart analytics that we expect of OLAP technologies are simply impossible.

Moreover, by making MDX into an industry standard, the language has become the lingua franca of the BI world. Virtually every BI client application and almost all OLAP servers have adopted MDX as their primary if not exclusive query language. In a short period of 5 years MDX has become to multidimensional databases what SQL is to relational databases. Can you imagine where the technology of relational databases would have been today without having a common query language?

Undoubtedly, Microsoft's entrance into the BI market at the end of 1998 with the release of OLAP Services in SQL Server 7.0 was the most influential event in the young BI industry. Starting from early 1997 when I and the small Plato development team assembled in Building 6 on the Redmond campus our mission was clear: "BI to the masses." We were tasked with creating a very powerful yet easy to use OLAP Server that could be distributed and assimilated on a mammoth scale and with unprecedented low prices.

While we looked at every release as the most important product release and had high expectations from every one of the versions we shipped, it is clear to us that bringing BI to the masses is a long journey that will take almost a decade to complete. We have made huge progress in the four years since we shipped the first release of OLAP Services. SQL Server 7.0 shook all previously known conventions of ease of use and prices in OLAP products. Analysis Services in SQL Server 2000 captured the market share leadership and became the most widely deployed OLAP Server ever. We are very pleased with the accomplishments so far but we know that we still have some way to go before BI is truly available to the masses.

We are now working on a major new release of Analysis Services under the code name Yukon and by the time these lines are published some of the dramatic innovations in the product will be known to the public. I view Yukon as the release that will bring BI servers very close to the point of realizing the "BI for the masses" vision. MDX stays a critical component of the platform and we are making a huge investment to make it even more powerful as well as much easier to use. The power of MDX is the power of analytics and what makes BI so important.

Mosha Pasumansky is the development lead for the MDX engine of Analysis Services in Microsoft. Mosha and I go a long way back to the days when we both worked for Panorama Software. From the moment I saw the code that he produced it was clear to me that Mosha is a prodigy and was one of the finest developers I have ever seen. Mosha took over from me the responsibility for the calculation engine of the Panorama OLAP product and he has been doing multidimensional calculations ever since.

When Microsoft acquired Panorama and the development team relocated to Redmond in January 1997 Mosha joined a very small task force that was responsible for creating Tensor, the code name for the OLE DB for OLAP specification. As part of that effort we defined the central component of the standard – the MDX language. MDX carries a lot of Mosha's genes in it. For years Mosha has been an authority on the practical usages of MDX to solve common business and analytical problems. Reading a book about MDX by Mosha is reading a book from the guy who knows everything there is to know about MDX.

Rob Zare joined the team for the late phases of SQL 2000 release. Astonishingly, Rob joined us with almost no background in computing and was hired to do some of the grunt testing work of the product; testing the user interface to ensure that it was not defective in any obvious way. Very quickly we discovered that Rob possesses explosive energy and a unique desire to excel. While Rob was doing the grunt work during the day, superbly at that, he moonlighted at building some impressive OLAP applications using the technology and writing up a bunch of product improvement suggestions. Very quickly Rob caught the attention of some of the senior members of the team who started mentoring him in the various aspects of product design. Soon afterwards, Rob got a double promotion and was reassigned to the team as a "program manager," a position that in Microsoft means a person that designs and writes the product specifications.

Rob's expertise is in building usable systems. He is responsible for some of the key aspects of the user interface in the Yukon release and he designed major portions of the MDX authoring tools. I am incredibly impressed with Rob's work and when Yukon is released I am sure that when you look at the outstanding product design you'll agree with me. Rob's passion is in making hard things easy and when you read this book I am sure you'll appreciate that MDX is presented in an easy to digest form thanks to Rob's work.

Mark Whitehorn is the professional author of the trio. He is the one that took Mosha's and Rob's ideas and formed them into an easy to read and entertaining text. I am sure you'll enjoy his write up as much as I did.

To truly understand modern Business Intelligence and to harness the power of the OLAP platform one must understand MDX, and this book, written by some of the creators of MDX, goes a long way in bringing the reader into the MDX way of thinking.

Amir Netz
Product Unit Manager
SQL Server – Analysis Services
Microsoft Corp.

Preface to the second edition

As this edition goes to print, SQL Server 2005 is looming large on the horizon. Well in advance of the release of SQL Server 2005, ProClarity (the company) has released version 6 of ProClarity (the tool). Version 6 has significant enhancements to the user interface that allow it to interact elegantly with both SQL Server 2000 and 2005. As the authors of this book, we wanted to be able to include the most recent version of ProClarity so we have gone right through the book, retaking the screen shots so that they match the new version of ProClarity. In addition, we have re-written the text where appropriate.

But there is more

As we say in the Introduction "... this isn't a complete reference to MDX; as the title tries to suggest, it is a fast track to learning MDX. It is designed to get you started quickly, and to give you the essential framework around which you can fill in the detail." We have been delighted with the reception of the first edition of the book and are determined to stick to our original ideal. So, apart from the changes described above, the first eighteen chapters of this book are very much as before.

Having said that... we couldn't resist adding 3 new chapters. Ah, but what to add? We could have simply expanded the range of functions that we introduce but that almost felt like cheating. The whole idea of the first edition of the book was not to list the entire function set but to show you how to use MDX to solve real business problems. So for the new material we decided to focus on a more advanced topic – recursion.

At its simplest, recursive code is code that calls itself. Within the realms of multi-dimensional data, recursion becomes a very powerful tool, as long as you understand what it is and how it works. The material we have added introduces recursion in a (hopefully) understandable way and shows you how to solve several business problems elegantly. We also take

a look at what is probably the most misunderstood MDX function of all, `NonEmptyCrossJoin`.

It could be argued that we have strayed from our ideal of writing an introductory book. But it could also be argued that by the time you have understood the first eighteen chapters, you're ready for some more advanced information. As a final defense of our actions, the cost of the book remains the same so even if you don't read the new material, you haven't lost anything. And, when you finally do need to know about recursion, the material is there waiting for you.

Introduction

This is where we try to convince you to buy the book and tell you what it tries to do and what it doesn't try to do. We also cover the housekeeping information such as introducing you to the sample files, pointing you to a web site for up-to-date information and generally setting the scene for the book. If you have already bought the book and/or know what it does, feel free to skip to Chapter 1 where the action starts. You can always come back later for the housekeeping information.

Why should you read this book?

OLAP (On-Line Analytical Processing) is an extremely potent tool and MDX (Multi-Dimensional eXpressions) is the key that unlocks the power of OLAP. If you have started to use Analysis Manager to create and/or use OLAP cubes then you'll rapidly reach the point where knowledge of MDX becomes useful, not to say essential. (OK, that's the major selling pitch over).

What is MDX?

Well, we suspect you have some kind of idea, otherwise the title of this book would not have attracted you, but for the record:

MDX is a language that allows you to query OLAP cubes in a way reminiscent of that in which SQL allows you to query relational databases. In addition MDX expressions (as they are called) can be used to add business logic to the cubes, to define simple and advanced security settings, to implement color coding for purposes of exception alerting, to create custom member roll-ups, custom level roll-ups, actions and so on – in other words, MDX is used almost everywhere in the design of effective OLAP cubes. If you build OLAP databases of any complexity then you are going to need MDX.