

TECHNOLOGY TRANSFER PRESENTS

# BILL INMON

---

## BUILDING THE TEXTUAL DATA WAREHOUSE

---

**MAY 10-11, 2010**  
RESIDENZA DI RIPETTA - VIA DI RIPETTA, 231  
ROME (ITALY)



[info@technologytransfer.it](mailto:info@technologytransfer.it)  
[www.technologytransfer.it](http://www.technologytransfer.it)

# BUILDING THE TEXTUAL DATA WAREHOUSE

## ABOUT THIS SEMINAR

For years corporate decisions have been made on the basis of the data found in transaction based systems. Transaction oriented data fits well with standard Database Management Systems because Database Management Systems structure data in a repetitive manner, where each occurrence of data has the same structure as each other occurrence of data in a table. But there is another viable and important source of data in the corporation. That source of data is the information found in the form of text. There are many forms of text in the corporation – emails, spreadsheets, contracts, warranties, medical and healthcare information, and so forth. Because text is not repetitive it does not fit easily and well with standard Database Management Systems.

But now there is textual ETL and the ability to build databases and Data Warehouses that contain textual information. When textual data is able to be transformed so that the text fits inside a standard Database Management System, whole new opportunities for analysis and Decision Making are created.

This two day lecture/workshop is about what is required to create the textual, unstructured Data Warehouse. The first day is lecture and the second day is a hands on workshop.

### WHO SHOULD ATTEND

This seminar/workshop is for people who are interested in the mechanics of taking text and producing an analytical database from that text. Data Architects, Business People, Project Managers, Technicians are all welcome in this class.

## OUTLINE

<p>DAY 1</p> <p><b>An Introduction To Unstructured Data</b></p> <p>This Powerpoint presentation shows what the issues of unstructured data are, what the possibilities, and what opportunities there are. The attendee is introduced to textual ETL and the creation of the unstructured database. The issues of integration are discussed amply.</p> <p><b>Issues of Textual Integration</b></p> <p>There are many issues in the integration of unstructured data and the transformation of unstructured data to. Some of the issues are:</p> <ul style="list-style-type: none"><li>• The issue of terminology</li><li>• The issue of logical sib divisions of data with a document</li><li>• The issues of clustering</li><li>• The issues of proximity</li><li>• The issues of filtering unnecessary data</li></ul> <p>These and many other issues are discussed in the context of reading in raw text and creating a viable analytical database.</p> <p><b>Forms of Text</b></p> <p>One of the major issues of textual ETL processing is that of the form of the raw text that is being processed. Some text is well formed – with proper spelling, nouns, verbs, prepositions and so forth. Other text is text that is in the form of comments, such as that found in doctor's notes. Yet other text is in an IM form. In fact there are many forms of text. In order to be effective textual ETL must account for all forms of text.</p>	<p><b>Spreadsheets</b></p> <p>A special case is that of reading Excell spreadsheets. The physical reading of Excell spreadsheets is not difficult. But the interpretation of the meaning of the many cells found in a spreadsheet is another story altogether. Yet Excell spreadsheets are another valid source of textual information.</p> <p><b>Diverse Indexes</b></p> <p>The results of textual ETL are diverse. Textual ETL produces many different forms of indexing, based on the type of document that is being processed. Some of the different types of indexes produced include:</p> <ul style="list-style-type: none"><li>• User specified word index</li><li>• Data fractured indexes</li><li>• Named value indexes</li><li>• Data fractured, semistructured index</li><li>• Named value, semi structured index</li><li>• Document Metadata index</li><li>• Data fractured/named value merged index</li></ul> <p>Each of these different kinds of indexes will be discussed. Their uses and their peculiarities will be addressed.</p>	<p>DAY 2</p> <p><b>Workshop</b></p> <p>On day 2 Textual ETL will be run producing a wide variety of Databases/Data Warehouses using many of the features of Textual ETL. The attendees will observe and participate in the transformation of text into a database ready for analytic processing.</p> <p>The workshop begins by examining some textual data. A strategy for capturing and organizing the text is discussed. Then the workshop continues with several types of processing that are done dynamically, under the purview of the attendees.</p> <p>Some of the types of processing that are done include:</p> <ul style="list-style-type: none"><li>• Document Metadata capture</li><li>• Document fracturing</li><li>• Named value indexing</li><li>• Simple indexing</li><li>• Semistructured indexing</li><li>• Merged indexing.</li></ul> <p>Depending on the textual data that has been selected, some or all of these kinds of indexes will be chosen and created.</p> <p>Many of the features of Textual ETL will be used during the workshop.</p> <p>At the end of the workshop the attendees may process some documents that they have brought to the workshop to see what the processing of textual ETL can do.</p>
--	---	--

## INFORMATION

<p><b>PARTICIPATION FEE</b></p> <p>€ 1200</p> <p>The fee includes all seminar documentation, luncheon and coffee breaks.</p> <p><b>VENUE</b></p> <p>Residenza di Ripetta Via di Ripetta, 231 Rome (Italy)</p> <p><b>SEMINAR TIMETABLE</b></p> <p>9.30 am - 1.00 pm 2.00 pm - 5.00 pm</p>	<p><b>HOW TO REGISTER</b></p> <p>You must send the registration form with the receipt of the payment to: <b>TECHNOLOGY TRANSFER S.r.l.</b> Piazza Cavour, 3 - 00193 Rome (Italy) Fax +39-06-6871102</p> <p><b>within</b> <b>April 26, 2010</b></p> <p><b>PAYMENT</b></p> <p>Wire transfer to: Technology Transfer S.r.l. Banca Intesa Sanpaolo S.p.A. Agenzia 6787 di Roma Iban Code: IT 34 Y 03069 05039 048890270110</p>	<p><b>GENERAL CONDITIONS</b></p> <p><b>GROUP DISCOUNT</b></p> <p>If a company registers 5 participants to the same seminar, it will pay only for 4. Those who benefit of this discount are not entitled to other discounts for the same seminar.</p> <p><b>EARLY REGISTRATION</b></p> <p>The participants who will register 30 days before the seminar are entitled to a 5% discount.</p> <p><b>CANCELLATION POLICY</b></p> <p>A full refund is given for any cancellation received more than 15 days before the seminar starts. Cancellations less than 15 days prior the event are liable for 50% of the fee. Cancellations less than one week prior to the event date will be liable for the full fee.</p> <p><b>CANCELLATION LIABILITY</b></p> <p>In the case of cancellation of an event for any reason, Technology Transfer's liability is limited to the return of the registration fee only.</p>
--	--	--

### BILL INMON

#### BUILDING THE TEXTUAL DATA WAREHOUSE

May 10-11, 2010  
Residenza di Ripetta  
Via di Ripetta, 231  
Rome (Italy)

Registration fee:  
€ 1200

*If registered participants are unable to attend, or in case of cancellation of the seminar, the general conditions mentioned before are applicable.*

first name .....

surname .....

job title .....

organisation .....

address .....

postcode .....

city .....

country .....

telephone .....

fax .....

e-mail .....



Stamp and signature

Send your registration form with the receipt of the payment to:  
**Technology Transfer S.r.l.**  
Piazza Cavour, 3 - 00193 Rome (Italy)  
Tel. +39-06-6832227 - Fax +39-06-6871102  
info@technologytransfer.it  
www.technologytransfer.it



## SPEAKER

**Bill Inmon** is a world-renowned expert, speaker and author on Data Warehousing, is widely recognized as the “*father of Data Warehousing*”. He is creator of the **Corporate Information Factory** and more recently, creator of the **Government Information Factory**. He has over 35 years of experience in database technology management and Data Warehouse design, and he is known globally for his seminars on developing Data Warehouses. He has been a keynote speaker for many major computing association and many industry conferences, seminars, and tradeshow. As an author, Mr. Inmon has written more than 650 articles on a variety of topics about building, using, and maintaining the Data Warehouse and the Corporate Information Factory. His works have been published in major computing journals including Data Management Review and The Business Intelligence Network where he continues to be a featured columnist. He has written 46 books, many of which have been translated into nine languages; one has sold over one-half million copies.