

# Winter 1996: Data Warehousing

## Speakers:

- **Ramon Barquin**, President and founder of the Data Warehousing Institute
- **Ian Paller**, President of the CIO Institute
- **Thomas A. DePasqule**, VP of Platinum Technology's Repository Group

## Synopsis of the Presentation

### 1. What is a data warehouse?

Definitions have evolved from Inmon's early definition: "A collection of data in support of management decision making" to Herb Edelstein's definition: "A data warehouse is the consolidation from multiple sources into a query database."

### 2. The Four components of warehouses

1. DSS front end (supply answers to the end users)
2. Source - Legacy, others (non legacy, external to organization)
3. Warehouse - Database Management Systems design may be centralized or distributed
4. Metadata - repository

### 3. Components Of Data Warehousing Process

1. Justification & System Integration
2. Moving & Staging Information
3. Information Access
4. Management

### 4. Mistakes to Avoid in Data Warehousing

1. Starting with the wrong sponsorship chain.
2. Setting expectations that you cannot meet and frustrating executives at the moment of truth
3. Engaging in Politically-Naive behavior.
4. Loading the warehouse with information "Just because it was available"
5. Believing that data warehousing database design is the same as transactional database design
6. Choosing a data warehousing manager who is technology-oriented rather than user-oriented
7. Focusing on traditional internal record-oriented data and ignoring the potential value of external data and of text, images, and - potentially - sound and video.
8. Delivering data with overlapping and confusing definitions
9. Believing the performance, capacity, and scalability promises.
10. Believing that once the data warehouse is up and running your problems

are finished.

11. Focusing on ad hoc data mining and periodic reporting.

5. Six patterns vendors will be pushing and users will be pulling

1. Web warehousing
2. Performance Boosts
3. Quality comes out of the closet
4. ROLAP fills OLAP niche
5. Mining for the masses
6. Warehouse enabled OLTP

6. Scope Control

1. What do I currently provide?
2. How much data is really needed?
3. Should I have summary files?
4. How do I convert existing queries?

7. Storage Goals Summary

1. Clear benefit focus
2. Definition quality
3. Timing consistency
4. Appreciation of reality
5. Search for "low-hanging fruit"

8. Additional Areas of interest that were discussed

1. Sizing of a data warehouse
2. Existing queries
3. Archiving strategies
4. Query response time management
5. Managing end-user queries

## **Speaker: Ramon Barquin**

Ramon Barquin is president and founder of the Data Warehousing Institute, as well as of Barquin and Associates, his own consulting firm. He specializes in developing information systems strategies, particularly data warehousing, for corporations. His presence in the information technology industry has spanned four decades and three continents. He had a long career with IBM covering both management and technical assignments, including overseas postings, and responsibilities in Asia and Latin America. Afterwards he served as President of the Washington Consulting Group, where he had direct oversight for the performance of several major federal information systems contracts. An electrical engineer and mathematician by training, he holds a Ph.D. degree from MIT. The author of over 75 technical and management publications, he has held faculty appointments at MIT, the Chinese University of Hong Kong, and the University of Maryland. An acknowledged expert in the field of data warehousing strategies, he has assisted the FAA, CIGNA, IBM, the

Department of the Treasury, and other organizations looking for guidance with data warehousing initiatives. Dr. Barquin is also the editor for the Prentice-Hall book series on data warehousing; and the co-editor of two coming volumes: *Planning and Designing the Data Warehouse* and *Building, Using, and Managing the Data Warehouse*.

### **Speaker: Allan Paller**

Allan Paller is director of education and research at the Data Warehousing Institute and also serves as president of the CIO Institute. He is the author of *Ten Mistakes to Avoid for Data Warehousing Managers* (TDWI 1995), *The EIS Book: Information Systems for Top Managers* (Irwin 1990), *How to Give the Best Presentation of your Life* (1982), and more than 100 articles. He has chaired the CIO Perspectives Conference sponsored by *CIO Magazine* and more than 40 other national and international conferences. His courses and briefings on effective technical presentations, on improving the CEO/CIO relationship, on the 10 mistakes to avoid in data warehousing, and on controlling the clandestine costs of client/server computing are consistently the top-rated programs in conferences around the world. His degrees are in information systems and engineering are from Cornell and the Massachusetts Institute of Technology.

### **Speaker: Thomas A. DePasquale**

Tom DePasquale is the Vice President of PLATINUM Technology's Repository Group. Since 1982, Tom has been a prominent advocate of Repository technology in support of Data Administration, Information Resource Management, Data Dictionaries, and Data Warehousing technology. He currently directs all activities of the PLATINUM Repository Labs, including product direction, international sales as well as the effort of the PLATINUM Information Management Consulting (PIMC) Group. Tom was one of the original founders and President of the Arlington-based RELTECH Products, Inc. which merged with PLATINUM in April of 1995. Tom spends a great deal of time advocating the benefits of repository to Fortune 1000 companies worldwide. Many of these organizations implement repository technology to solve a combination of five basic needs: Legacy System documentation, CASE Model Management, Data Warehousing, Migration to Client/Server, and traditional Data Dictionary support.