Dynamic Data Warehousing

Vince Leat
ASEAN SW Group

August 2007
Discussion

- Introduction
- Dynamic Data Warehousing
  - Data Partitioning
  - Workload Control
  - Compression
  - Data Mining
  - Industry Models
- Balanced Warehouse
Top of Mind Issues for the World’s Leading CEOs

<table>
<thead>
<tr>
<th>CEO Challenges</th>
<th>What’s Left?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Competitors offer <strong>similar products</strong></td>
<td>Execute with <strong>maximum efficiency and effectiveness</strong></td>
</tr>
<tr>
<td>Competitors use <strong>similar technology</strong> and proprietary technologies are <strong>quickly copied</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Breakthrough innovation</strong> in products or services is <strong>increasingly difficult</strong></td>
<td><strong>Make the smartest business decisions</strong> possible</td>
</tr>
<tr>
<td><strong>The World is Flat</strong> – losing geographic &amp; regulatory advantages</td>
<td></td>
</tr>
</tbody>
</table>
An Information Based Strategy

“In God we trust; all others bring data”
- W. Edwards Deming
The Next Generation of Business Intelligence

*Insightful, Relevant Information When and Where it’s Needed*

- **Information On Demand**
  - to Optimize Real-Time Processes

- **OLAP & Data Mining**
  - to Understand Why and Recommend Future Action

- **Query & Reporting**
  - to Understand What Happened

- **Help Solve Crimes by Delivering Suspect List to Detectives Arriving at the Crime Scene**

- **Optimizing Police Force Deployments**

- **Crime Rate Reports**
Using Information to Compete
Driving Real Business Results in Financial Services

Using Analytics to Compete

- Identify most profitable customers and minimize risk to maximize risk-value reward
- Adjust CD rates & rollover incentives to attract and retain best customers at lowest price

Business Results

- Increased customer retention for savings accounts by 87%
- Lowered customer acquisition costs by 83%
- Drove 25% increase in revenue per customer account over first 3 years
Using Information to Compete

*Driving Real Business Results in Retail & Distribution*

### Using Analytics to Compete

- Maintaining lowest possible inventory while insuring availability
- Identify most loyal customers and maximize cross-selling

### Business Results

- Saved $1.2 billion in inventory costs over 5 years
- Increased market share from 36% to 43%
- Drove same store sales gains in 23 of 24 quarters over 6 year period
Dynamic Warehousing
A New Approach to Leveraging Information

**Information On Demand**
to Optimize Real-Time Processes

**Dynamic Warehousing**

**Dynamic Warehousing Requires:**
1. Real-time access – in context
2. Analytics – as part of a business process
3. Unstructured information – extracted knowledge
4. Extended infrastructure – tightly integrated

**Query & Reporting**
to Understand What Happened

**OLAP & Data Mining**
to Understand Why and Recommend Future Action
Dynamic warehousing
*Extending beyond the warehouse to enable information on demand*
More Examples of Dynamic Warehousing in Action

*Enabling Information On Demand for Business Advantage*

<table>
<thead>
<tr>
<th>Traditional warehousing</th>
<th>Dynamic warehousing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insurance fraud analysis and reporting</td>
<td>Identifying potentially fraudulent claims prior to approval and payment</td>
</tr>
<tr>
<td>Reporting on customer issues</td>
<td>Identifying possible related issues, churn risk and cross-sell opportunities while engaged with the customer</td>
</tr>
<tr>
<td>Historical sales analysis and reporting</td>
<td>Discovering relevant customer information to identify cross sell opportunities and improve negotiating position at the point of sale</td>
</tr>
</tbody>
</table>

*Transforms healthcare*

*Transforms customer service*

*Transforms sales effectiveness*
What are the challenges for traditional warehouses?

- Leverage ALL information, including unstructured
- Address expanding needs for analytics
- Increasing types of applications lead to varying service level demands
- Not just for traditional query and reporting

Increasingly mixed workload environments and the constantly changing needs of different business constituents require more dynamic warehousing capabilities.
IBM provides more than just a warehouse

**DB2 Warehouse provides extended capabilities and value**

### Embedded analytics
**Inline and as a Service**
- Multidimensional analysis
- Data mining and visualization

### Beyond traditional structured data
- Generate and leverage knowledge from unstructured information

---

**IBM DB2 Warehouse**

---

**Best of Both Worlds Architecture**

- Benefits of an embedded data server
  - Optimized for
  - High availability and reliability
  - Scalable, secure and auditable

- Beyond traditional structured data
  - Advanced data partitioning
  - Workload management
  - Better disk utilization
  - Query speed improvement

---

© 2007 IBM Corporation
How IBM Enables Dynamic Warehousing

*Integrated offerings to enable information on demand*

![Diagram of IBM offerings](image)

- **Search & Text Analytics**
  - OmniFind™ Analytics Edition

- **Enterprise Data Modeling**
  - Rational® Data Architect

- **Information Integration**
  - Information Server

- **Process Mgmt**
  - FileNet BPM
  - WebSphere BPM

- **MDM**
  - WS Customer Center
  - WS Product Center

- **Industry Perspective**
  - IBM industry data models

- **SOA Infrastructure**

- **IBM DB2 Warehouse**

- **IBM Global Services and Business Partners**
Warehousing strategic pillars

*Guiding principles for innovation*

**Simplicity**
- Easy to deploy and integrate
- Easy to use
- Easy to manage
- Easy to start and grow as needed

**Reliability & Performance**
- Reliable access to information
- Highly available
- Real-time performance
- Maximized resource efficiency

**Actionable Insight**
- Beyond traditional capabilities
- Further leverage information
- Extended business insight
- Support broader usage
IBM Warehouse software
A complete, integrated platform

Embedded analytics
- Data mining, modeling and scoring
- Visualization and in-line analytics

Performance optimization
- Data partitioning
- Workload control
- Deep compression

Data movement and transformation

Database management

DB2 Warehouse
Warehousing Made Simple

Common Eclipse Based Design Studio for All Administration

Data Warehouse Project

Enterprise Schema

Data Flow

Control Flow

SQL Generated From Data Flow

Generated From Data

Flow

Data

Warehouse

Project

Simplicity
Data Partitioning

- Partition a database within a single server or across a cluster of servers
  - Scale to support very large data sets
  - Minimize impact of complex workloads
  - Provide increased parallelism for administration tasks
Workload Control

“Query Traffic Corp”
- Prioritize queries across applications
- Ensure operational processes are serviced first
- Minimize impact of complex queries or batch processes

“Accountant”
- Keep a lid on “cost”
- Real-time & historical query-execution statistics
- Reports on:
  - data accessed most frequently
  - data NOT being accessed
  - users or groups generating most workload
Deep Compression
Reduced storage costs – better disk utilization – faster queries

Complex Queries Comparison
(Lower is better)

Compression Leads to Faster Queries!
Data Mining

*Enhancing Business Insight with Predictive Analytics*

- **Associations**
  - Which item affinities (“rules”) are in my data?
  - \([\text{Beer} \Rightarrow \text{Diapers}] \ldots\text{single transaction}\)

- **Sequences**
  - Which sequential patterns are in my data?
  - \([\text{Love}] \Rightarrow [\text{Marriage}] \Rightarrow [\text{Baby Products}] \ldots\text{sequential transactions}\)

- **Clustering**
  - Which interesting groups are in my data?
  - \(\ldots\text{customer profiles, store profiles}\)

- **Classification**
  - How to predict categorical values in my data?
  - \(\ldots\text{will the patient be cured, harmed, or unaffected by this treatment?}\)

- **Prediction**
  - How to predict numerical values in my data?
  - \(\ldots\text{how likely a customer will respond to the promotion}\)
  - \(\ldots\text{how much will each customer spend this year?}\)

- **“Easy Mining” algorithms**

- **Score data directly in DB2, scalable and real time**
Embedded mining with integrated tools

*Seamless integration of analytics capabilities*

- Filter required data directly in the warehouse
- Get subset of data for analysis
- Drag-and-drop interface
  - Seamlessly add specific analytics and mining operations
Deliver inline visualization and analytics

*Embedded analytics capabilities*

Can be embedded directly into applications and Web pages

Out-of-the-box visualization tools
# Industry data models

_Leverage industry best practices for faster time to market_

## Banking
*(Banking Data Warehouse)*
- Profitability
- Relationship marketing
- Risk management
- Asset and liability management
- Compliance

## Financial Markets
*(Financial Markets Data Warehouse)*
- Risk management
- Asset and liability management
- Compliance

## Health Plan
*(Health Plan Data Warehouse)*
- Claims
- Medical management
- Provider and network
- Sales, marketing and membership
- Financials

## Insurance
*(Insurance Information Warehouse)*
- Customer centricity
- Claims
- Intermediary performance
- Compliance
- Risk management

## Retail
*(Retail Data Warehouse)*
- Customer centricity
- Merchandising management
- Store operations and product management
- Supply chain management
- Compliance

## Telco
*(Telecommunications Data Warehouse)*
- Churn management
- Relationship management and segmentation
- Sales and marketing
- Service quality and product lifecycle
- Usage profile

---

*Over 400 Customers!*

*New Offering!*
Introducing IBM Balanced Warehouse™
A fast track to warehousing

Balanced Configuration Modules
Preconfigured, pretested allocation of software, storage and hardware to support a specified combination of function and scale

Simplicity
- Predefined configurations for reduced complexity
- One number to contact for complete solution support

Flexibility for growth
- Add BCU’s to address increasing demands
- Multiple on-ramps for different needs
- Reliable, nonproprietary hardware for reusability

Optimized performance
- Preconfigured and certified for guaranteed performance
- Based on best practices for reduced risk

Better than an appliance
IBM Balanced Warehouse Offerings

**C-Class**
- Out of the Box Warehousing Solution for SMB Customers
  - (includes out-of-the-box BI tools)
  - Linux on System x

**D-Class**
- Departmental Data Marts and Small to Mid-Size Data Warehouses
  - Linux on System x

**E-Class**
- Large Enterprise Data Warehouses
  - AIX on System p

<table>
<thead>
<tr>
<th></th>
<th>C-Class</th>
<th>D-Class</th>
<th>E-Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>Starting at $15K</td>
<td>Starting at $275K</td>
<td>Starting at $850K</td>
</tr>
<tr>
<td>Storage</td>
<td>Scales to 1 TB</td>
<td>1TB to 5TB (target)</td>
<td>4TB and Up</td>
</tr>
<tr>
<td>Users</td>
<td>Up to 200 Users</td>
<td>Modular Scalability</td>
<td>Unlimited, Modular Scalability</td>
</tr>
<tr>
<td>Hardware &amp; Storage</td>
<td>Affordable Hardware &amp; Storage</td>
<td>Mid-Range Hardware &amp; Storage</td>
<td>High End Hardware &amp; Storage</td>
</tr>
</tbody>
</table>
IBM is the Leading Provider of Warehousing
The Industry Leaders Use DB2 for Warehousing

11 of the top 12 banks
7 of the top 8 auto manufacturers
5 of the top 6 insurance companies
4 of the top 6 general merchandisers
4 of the top 5 specialty retailers
3 of the top 4 food & drug stores
IBM enables dynamic warehousing

*Delivering greater value from information*

- More dynamic and balanced approach to warehousing is key
- Broad set of capabilities beyond the warehouse required
- IBM provides the most comprehensive platform to address these needs
Copyright information

© Copyright IBM Corporation 2007

IBM Corporation
Software Group
Route 100
Somers, NY 10589
U.S.A.

Produced in the United States of America
03-07
All Rights Reserved.

DB2, IBM, the IBM logo, OmniFind, Rational and System z are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries or both.

Other company, product and service names may be trademarks or registered trademarks or service marks of others. The information contained in this documentation is provided for informational purposes only. While efforts were made to verify the completeness and accuracy of the information contained in this documentation, it is provided “as is” without warranty of any kind, express or implied. In addition, this information is based on IBM’s current product plans and strategy, which are subject to change by IBM without notice. IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, this documentation or any other documentation. Nothing contained in this documentation is intended to, nor shall have the effect of, creating any warranties or representations from IBM (or its suppliers or licensors), or altering the terms and conditions of the applicable license agreement governing the use of IBM software.