Data Warehousing
Business Intelligence

An Introduction
About

• Robert C. Cain
• Southern Nuclear since 2005
• 10 years as a consultant in the B’ham Market
• Wide range of .Net applications, ASP & Win
• SQL Server 2005 Data Warehouse
• http://arcanecode.com
• http://twitter.com/arcanecode
• http://altechevents.com
What is a Data Warehouse

- A giant storehouse for your data
- ALL of your data
- Aggregation of data from multiple systems
What is Business Intelligence

• Leveraging data you already have
• Examining the data in your warehouse to look for:
  • Aggregations
  • Trends
  • Correlations (Data Mining)
Why Have a Data Warehouse?

- Combine data from multiple systems and resolve inconsistencies between those systems
- Make reporting easier
- Reduce the load on production systems
- Provide for long term storage of data
- Provide consistency among system transitions
Some More Reasons for a Data Warehouse

• Make the data available for analysis
• Ability to apply advanced data mining tools
• To extract further value from the data you already own

Business Intelligence
Business Intelligence is HOT

• According to Computerworld, BI is the 5th hottest IT Skill for 2009
• Dice.com over 2,800 job openings
What's wrong with reporting from a Transactional System?

- OLTP – On Line Transaction Processing
- Designed for working with single record at a time.
- Data is highly “normalized”, i.e. duplicate values have been removed.
- Getting all data for a record can involve many table joins
- Can be quite confusing for ‘ad-hoc’ reporting
- Can also be slow, having an impact on the OLTP system
What’s different about a Data Warehouse?

• Data Warehouses typically use a design called OLAP
• On-Line Analytical Processing
• Data is de-normalized into structures easier to work with.
• Number of tables are reduced, reducing number of joins and increasing simplicity
• Often a Star Schema or Snowflake Schema
Snowflake Schema
Types of Tables in a Warehouse

- Dimensions
- Facts
- Both require the concept of Surrogate Keys
- A new key, typically some type of INT, that is used in place of any other key as the Primary Key
Reasons for Surrogate Keys

- Preserve data in case of source system change
- Combine data from multiple sources into a single table
- Source System keys can be multi-column and complex, slowing response time
- Often the key is not needed for many data warehousing functions such as aggregations
Dimensions

- Dimensions hold the values that describe facts
- “Look Up Values”
- Some examples: Time, Geography, Employees, Products, Customers
- When a Dimension can change over time, it’s known as a Slowly Changing Dimension
- Three types of Dimensions: Type 1, 2, & 3
Type 1 Dimension

- When a dimensions value is updated, the old one is simply overwritten

<table>
<thead>
<tr>
<th>ID</th>
<th>Last</th>
<th>First</th>
</tr>
</thead>
<tbody>
<tr>
<td>1234</td>
<td>McGillicutty</td>
<td>Hortence</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>Last</th>
<th>First</th>
</tr>
</thead>
<tbody>
<tr>
<td>1234</td>
<td>Hollywoger</td>
<td>Hortence</td>
</tr>
</tbody>
</table>
Type 2 Dimension

- When a dimension is changed, a new record is inserted and old one dated

<table>
<thead>
<tr>
<th>ID</th>
<th>Last</th>
<th>First</th>
<th>FromDate</th>
<th>ThruDate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1234</td>
<td>McGillicuty</td>
<td>Hortence</td>
<td>12/1/1998</td>
<td>&lt;NULL&gt;</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>Last</th>
<th>First</th>
<th>FromDate</th>
<th>ThruDate</th>
</tr>
</thead>
<tbody>
<tr>
<td>2468</td>
<td>Hollywoger</td>
<td>Hortence</td>
<td>7/6/2008</td>
<td>&lt;NULL&gt;</td>
</tr>
</tbody>
</table>
Type 3 – Just Say NO

• When a dimensions value is updated, a new column is added

<table>
<thead>
<tr>
<th>ID</th>
<th>Last1</th>
<th>First</th>
</tr>
</thead>
<tbody>
<tr>
<td>1234</td>
<td>McGillicutty</td>
<td>Hortence</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>Last1</th>
<th>Last2</th>
<th>First</th>
</tr>
</thead>
<tbody>
<tr>
<td>1234</td>
<td>Hollywoger</td>
<td>McGillicutty</td>
<td>Hortence</td>
</tr>
</tbody>
</table>

• Almost never used
When pulling in data from multiple systems, you often have to reconcile different primary keys.

This process is known as conforming your dimensions.

<table>
<thead>
<tr>
<th>ID</th>
<th>Product</th>
<th>InventoryID</th>
<th>PurchasingID</th>
<th>WorkMgtID</th>
</tr>
</thead>
<tbody>
<tr>
<td>9876</td>
<td>Widget</td>
<td>459684932</td>
<td>Wid45968</td>
<td>602X56VV1</td>
</tr>
</tbody>
</table>
Fact Tables

• A Fact marks an event, a discrete happening in time
• Facts usually hold numeric measures and/or links to dimensions ID: SoldBy, SoldTo, Product Qty, SaleAmt, SaleDate
Fact Table Example - Sales

<table>
<thead>
<tr>
<th>ID</th>
<th>SoldByID</th>
<th>SoldToID</th>
<th>ProductID</th>
<th>Qty</th>
<th>SaleAmt</th>
<th>SaleDate</th>
</tr>
</thead>
<tbody>
<tr>
<td>3456</td>
<td>1234</td>
<td>6789</td>
<td>987</td>
<td>3</td>
<td>156.00</td>
<td>7/17/2009</td>
</tr>
</tbody>
</table>

- Employee Dimension
- Customer Dimension
- Product Dimension
Getting Data Into A Warehouse

- ETL
- Extract
- Transform
- Load
- SSIS – SQL Server Integration Services
Getting Data Out of Your Warehouse

- SSRS – SQL Server Reporting Services
- SSAS – SQL Server Analysis Services
• Key Performance Indicators
• Dashboards
• Quick, at a glance indicator of system health

<table>
<thead>
<tr>
<th>Region</th>
<th>Sales (USD)</th>
<th>Trending</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>US</td>
<td>482m</td>
<td>🔄</td>
<td>🟢</td>
</tr>
<tr>
<td>Europe</td>
<td>399m</td>
<td>🔄</td>
<td>🟥</td>
</tr>
<tr>
<td>Asia</td>
<td>123m</td>
<td>🔄</td>
<td>🟢</td>
</tr>
<tr>
<td>South America</td>
<td>225m</td>
<td>🔄</td>
<td>🟢</td>
</tr>
</tbody>
</table>
Warehousing Methodologies

- Inmon – Bill Inmon - Top down
- Kimball – Ralph Kimball - Bottom up
The Data Warehouse Toolkit by the Kimball Group

Resources

Smart Business Intelligence Solutions with Microsoft SQL Server 2008

http://www.amazon.com/Business-Intelligence-Solutions-Microsoft%C2%AE-PRO-Developer/dp/0735625808/ref=sr_1_1?ie=UTF8&s=books&qid=1239580654&sr=1-1
Resources

Programming Microsoft SQL Server 2008

• http://www.amazon.com/Programming-Microsoft-Server-2008-PRO-Developer/dp/0735625999/ref=sr_1_1?ie=UTF8&s=books&qid=1239580376&sr=1-1
Resources - SSIS

- **Erik Veerman Books**
  - [http://www.amazon.com/Professional-Microsoft-Integration-Services-Programmer/dp/0470247959/ref=sr_1_1?ie=UTF8&s=books&qid=1239833324&sr=8-1](http://www.amazon.com/Professional-Microsoft-Integration-Services-Programmer/dp/0470247959/ref=sr_1_1?ie=UTF8&s=books&qid=1239833324&sr=8-1)
Thanks Again!

- Questions?
- All material available at http://arcanecode.com
- arcanecode@gmail.com