Integrating an OLAP Data Cube with ArcMap

Lt. (j.g.) William Winner, NOAA
Deb Agarwal, UC Berkeley
Catharine van Ingen, Microsoft
Project Origins

- Began at UC Berkeley
  - Proposal to NSF for a watershed data synthesis center
  - Proposal to Microsoft to create a demonstration centered around California water issues (2006)
Original Objectives

• Database requirements
  – Needed large storage capacity
  – Needed ability to merge data from different agencies in different formats
  – Needed to handle different types of data
  – Needed the ability to perform analyses quickly
What is a Data Cube?

- An N-dimensional database
  - Traditional DBs can do this too, but...
    - Data must be flattened
    - Done using Aggregations, Group By’s and Cross Tabs
What is a Data Cube?


<table>
<thead>
<tr>
<th>Model Year</th>
<th>Color</th>
<th>Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chevy 1990</td>
<td>red</td>
<td>5</td>
</tr>
<tr>
<td>Chevy 1990</td>
<td>white</td>
<td>87</td>
</tr>
<tr>
<td>Chevy 1990</td>
<td>blue</td>
<td>62</td>
</tr>
<tr>
<td>Chevy 1991</td>
<td>red</td>
<td>54</td>
</tr>
<tr>
<td>Chevy 1991</td>
<td>white</td>
<td>95</td>
</tr>
<tr>
<td>Chevy 1991</td>
<td>blue</td>
<td>49</td>
</tr>
<tr>
<td>Chevy 1992</td>
<td>red</td>
<td>31</td>
</tr>
<tr>
<td>Chevy 1992</td>
<td>white</td>
<td>54</td>
</tr>
<tr>
<td>Chevy 1992</td>
<td>blue</td>
<td>71</td>
</tr>
<tr>
<td>Ford 1990</td>
<td>red</td>
<td>64</td>
</tr>
<tr>
<td>Ford 1990</td>
<td>white</td>
<td>62</td>
</tr>
<tr>
<td>Ford 1990</td>
<td>blue</td>
<td>63</td>
</tr>
<tr>
<td>Ford 1991</td>
<td>red</td>
<td>52</td>
</tr>
<tr>
<td>Ford 1991</td>
<td>white</td>
<td>9</td>
</tr>
<tr>
<td>Ford 1991</td>
<td>blue</td>
<td>55</td>
</tr>
<tr>
<td>Ford 1992</td>
<td>red</td>
<td>27</td>
</tr>
<tr>
<td>Ford 1992</td>
<td>white</td>
<td>62</td>
</tr>
<tr>
<td>Ford 1992</td>
<td>blue</td>
<td>39</td>
</tr>
</tbody>
</table>
What is a Data Cube?

What is a Data Cube?

How is the Cube stored?

- The Cube is a computed product from a series of relational tables.
- Our Cube is based on a Context-Dependent Snowflake Schema.
Snowflake Schema

Snowflake Schema
Context-Dependent Snowflake
NMFS Involvement

- Came aboard at the end of 2008
- Saw the potential to use this to help inform Salmonid Recovery Plans
- Saw the potential to include a spatial component with the data
Toolbar Requirements

- Allow for spatial queries
- Flatten data to fit within a shapefile’s attribute table
- Provide some way to visually move through the data
How did we do it?

• The toolbar was written using Visual Studio 2008 along with the ArcGIS API’s.
• Also included project management and storage
• Written with just shy of 10,000 lines of designer- and user-generated code
Data Cube Utility Toolbar

View Sites  Site Selection Wizard
Data Cube Utility Toolbar
Data Cube Utility Toolbar
Data Cube Utility Toolbar
Data Cube Utility Toolbar

Site Selection Wizard

- Select All Types
- NOAA Sites
- USGS Sites
- SEC Sites

Step 1 | Step 2 | Step 3

Cancel Back Next Proceed to DataCube
Data Cube Utility Toolbar
Data Cube Utility Toolbar
Data Cube Utility Toolbar
Data Cube Utility Toolbar
Future Goals

- Integrate spatial attributes directly into the cube
- Improve post-query speed
- Allow for other cubes
- Incorporate other data types