Abstract

What’s going on in your physical data model? How many people can or will update it to match the reality of what’s going on in your databases? Who decides what goes into the physical model?

In this presentation we discuss 10 physical data modeling mistakes that cost you dearly. Will your physical design lead to performance snags, development delays, bugs and weakening of professional respect?

Data Architects are often tasked to prepare first cut physical data models, yet these skills usually overlap those of DBAs and Developers and this overlap can lead to contention, confusion, and complacency. With this presentation, you’ll learn about the 10 blunders, how to find them, plus 10 tips on how to avoid them.
Karen López is a principal consultant at InfoAdvisors. She specializes in the practical application of data management principles. Karen is also the ListMistress and moderator of the InfoAdvisors Discussion Groups at www.infoadvisors.com and dm-discuss.

About this Presentation
Your opinion counts
Contributions are required
Ask Questions at any time

@datachick, with hashtag #10Blunders
Agenda

- The Problem
- Blunders, FAILs, D’Oh’s, Faults and WTHs?
- 10 Tips

POLL: Who Are You?
POLL: Physical Model Much?

The Problem & The Solution

© 2011 InfoAdvisors
Assuming Numbers are Numbers

D’OH!

Leading Zeros
Non-Numeric
Special Characters
Sorting Issues
Externally Managed Numbers
Add More Examples.

<table>
<thead>
<tr>
<th>Numbers that Aren't Numbers</th>
<th>Dates/Times that Aren't Dates/Times</th>
<th>Texts that Aren't (Really) Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Security Numbers</td>
<td>Birthdays</td>
<td>Format embedded values</td>
</tr>
<tr>
<td>Vehicle Identification Numbers</td>
<td>Intervals</td>
<td>???</td>
</tr>
</tbody>
</table>

Telephone Numbers

Account Numbers
Some Additional NTANs

<table>
<thead>
<tr>
<th>Numbers that Aren’t Numbers</th>
<th>Dates/Times that Aren’t Dates/Times</th>
<th>Texts that Aren’t (Really) Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Security Numbers</td>
<td>Birthdays</td>
<td>Format embedded values</td>
</tr>
<tr>
<td>Vehicle Identification Numbers</td>
<td>Intervals</td>
<td>???</td>
</tr>
<tr>
<td>Telephone Numbers</td>
<td>MS SQL Server Timestamps</td>
<td></td>
</tr>
<tr>
<td>Account Numbers</td>
<td>Lengths of time</td>
<td></td>
</tr>
<tr>
<td>ZIP Codes</td>
<td>Days</td>
<td></td>
</tr>
<tr>
<td>UPC/GTIN/EAN/Barcodes</td>
<td>Chris Date</td>
<td></td>
</tr>
</tbody>
</table>

Avoid it

1. Use a numeric-ish datatype when there’s math involved.
2. Do your research
3. If it’s externally managed data, its format or composition could change at any time. Anticipate that.
4. Anticipate leading zeros.
5. Anticipate significant formatting, characters & other tricks.
Choosing an Silly Long Primary Key

32 + 8 + 32 + 32 + (0 to 22) = A really big key that only a Data Architect could love
Avoid it

- Establish Primary Key standards and styles
- Smaller PKs lead to smaller indexes and smaller databases. Take advantage of that
- Choose PKs that do not change when data changes

DEMO
Turning off RI for Development

DEMO
Avoid it

Educate team on the risks & nearly universal outcome of turning off RI
Generate RI in every script

Develop test data
Develop scripts for loading data
Ensure team has easy access to Data Models

Using the Defaults
Defaults <> Best Practices
Product Defaults aren’t YOUR defaults
Defaults might even be...faulty.

DEMO
Avoid it

Create a DA Test database...and use it.
Test Generation Options to Test DB with Data
Generate Test Scripts, Varying Options
Review options with DBAs and Developers.
Choose Defaults, don’t default them
Save your Default Set

Using GUIDs Where They Don’t Add Value
GUID

A globally unique identifier or GUID ...is a unique reference number used as an identifier in computer software. The term GUID also is used for Microsoft's implementation of the Universally Unique Identifier (UUID) standard.

The value of a GUID is represented as a 32-character hexadecimal string, such as {21EC2020-3AEA-1069-A2DD-08002B30309D}, and is usually stored as a 128-bit integer. The total number of unique keys is $2^{128}$ or $3.4 \times 10^{38}$ — roughly 2 trillion per cubic millimeter of the entire volume of the Earth. This number is so large that the probability of the same number being generated twice is extremely small.

Applying a Surrogate Key
Avoid it

- Establish Primary Key standards and styles
- Smaller PKs lead to smaller indexes and smaller databases. Take advantage of that.
- Don’t forget the Business Key
- Choose PKs that do not change when data changes
- Understand how Identity columns work

Avoid It

1. Understand Datatype Sizes
2. Understand Database Internal File Structures
3. Determine Data Volume
4. Determine Data Growth Volumes
5. Balance What-if and What-Will-Be
Using Identity Property or Identifier Incorrectly

DEMO
Failing to Compare

DEMO
Silly Naming Standards

SHOW AND TELL
Skipping Training

DEMO
OTHER BLUNDELS?

Other Blunders

- Completely separate logical model
- Failing to take into account whole environment.
- Not testing your design
- Ignoring reference data design
- Hand generating scripts
- Failing to report issues
- Denormalizing too soon
- Just-in-case-itis
- Making performance more important than integrity
10 Tips for Avoiding Physical Modeling Blunders

1. Understand cost, benefit and risk
2. Get formal training on target technologies
3. Work near your DBAs and Developers
4. Ask DBAs about trade-offs, not just solutions
5. Build portfolio of performance vs. integrity trade-offs
10 Tips for Avoiding Physical Modeling Blunders

6. Profile Source Data
7. Build Test Databases, with Data
8. Test Scripts
9. Compare, Compare, Compare...then Compare some More.

Summary

1. Don’t assume that numbers are….numbers
2. Choose the right primary key
3. Apply surrogate keys correctly
4. Keep RI turned on
5. Architect, don’t default
6. Keys are blunder-prone
7. Training is required. Don’t pass up any opportunity for training.
Thank You!

Feedback, questions, comments are always appreciated.

http://www.speakerrate.com/karenlopez
http://blog.infoadvisors.com