Data Modeling Contentious Issues:

Karen Lopez
InfoAdvisors, Inc.
www.infoadvisors.com

Karen López, ISP

• Karen has 20+ years of data an information architecture experience on large, multi-project programs.
• She is a frequent speaker on data modeling, data-driven methodologies and pattern data models.
• She is a Microsoft MVP, SQL Server

• She wants you to love your data.

@datachick
facebook.com/lopezk
linkedin.com/in/karenlopez
About this Presentation

We will be using an interactive format - you will be participating in informal polls about data modeling issues and best practices.

This is not an introductory presentation - a good knowledge of data modeling issues will be assumed.

Why Contention?

...because people are contentious!
InfoAdvisors’ Discussion Groups

- E-mail, web, RSS and newsgroup based discussion group
- Data Modeling, Frameworks, Tools Groups
- Over 3000 subscribers
- Moderated
- No Charge
- www.infoadvisors.com

E-mail
RSS Feeds

Contentious DM Issues

Agenda

• Contentious Issues
  • Background
  • Discussion Quotes
  • Poll
  • Results & Analysis

• Resources
Contentious Issues

Near-religious discussions and debates

- People rarely change their minds based on this discussion
- The most successful discussions are ones where both sides learn something new about the other viewpoint.

Do we need Classwords?

- A traditional naming convention
- Usually means there’s a standard classword list
- Some tools can check for standards compliance
- Examples: Date, Amount, Count, Quantity
Attribute Names

- Customer First Name
- Customer First Name
- Order Effective Date
- E-mail Address
- Backordered Specialty Order Item ID
- Item ID
- Item

...not the same as

- Domains
- Data Types
- UDTs
Class words

Most DA's have very strong opinions on class words. Usually, the DA with the strongest opinion wins this debate. Every company where I've worked has a different set of class words based on their business needs.

The only point I ever bring up in addressing class words is not to have too many of them. The more you have, the more difficult the class words are to manage.

- Walter

Class words

In my opinion it all boils down to 3 basic things:

Exception, common sense and compromise. You can employ all three and still have a very strong set of standards. Standards must be workable to be affective. That means they must have some flexibility or you will never get buy in from those you want to use them.

- Garry
Class words

I have come to see that class words are metadata, and that what they represent, along with other meta-attributes of data, should be kept in some sort of metadata repository. Trying to add metadata as part of a field's name seems to me (1) to waste about 3 or 4 bytes that could profitably be used otherwise, (2) be a constant source of contention between folks, and (3) in the end add very little to what one already knows from data type definitions.

- Joseph

Our Process...

...vote early and vote often!
Vote!

I post a question

I give the answers a range.

You put the sticky in the area for your vote

We debate the answer and the results

Vote! LDM & PDM

1

LDM

2

3

4

5

PDM
Vote! General

1  2  3  4  5

LDM

PDM

Ready?
Vote: Class words

Does a good LDM / PDM data model use classwords?

1. Always
2. 
3. 
4. 
5. Classwords are so obsolete....

Class words

TRAP!

You may find yourself debating a topic because there is no common definition of a model or object, not because there is a real disagreement.
Keys - Natural or Surrogate

- Just what you call them may show your true colors....
  - Surrogate, Non-intelligent, Unnatural, Dataless or Meaningless
  - Natural, Intelligent, Normal, Cluttered
Keys - Natural or Surrogate - Vote

What best describes your approach in an LDM / PDM?

1. Surrogate Keys? We don’t need surrogate keys - we have natural identifiers.
2. 
3. 
4. 
5. Every entity deserves its own surrogate key.

Surrogate Keys

TIP!
Every decision is ultimately made based on cost, benefit, and risk. Be prepared to analyze all three.
When To Use Varchar?

- Variable in length
- Extra storage
- Some trick for managing
- Varchar(1)?
- Varchar(2)?
- Varchar(8)?

Varchar - Vote

What makes sense?

1. Whatever anyone wants (1, 2, 3 is fine)
2.
3.
4.
5. There is a minimum number where varchar should not be used
Process Models

• Workflows, DFDs, Use Cases, etc.
• Sometimes address data issues
• Off again, on again
• May be collaborative, may not

Process Models

• Business Analysts
• Developers
• UMLers
• Enterprise Architects
• Process Analyst/Modeler
Vote: Process Models

Which should drive efforts?

1. Data Models
2.
3.
4.
5. Process Models

Relationships

- **CustomerID**
  - Name
- **EmployeeID**
  - Name
  - BirthDate
- **OrderNumber**
- **OrderDate**

- CustomerID approves EmployeeID
- CustomerID takes OrderNumber
- EmployeeID takes OrderDate
Foreign Keys

- CustomerID
  - Name

- EmployeeID
  - Name
  - BirthDate

- OrderNumber
- Takes
- Approves

- Takes
- ApprovedByEmployeeID (FK)
- CustomerID (FK)
- OrderDate
- TakenByEmployeeID (FK)

Relationships & Foreign Keys

- FKs: results of Relationships
- Migrated attribute may be role named
- Relationships may be named…or not.
Vote: Relationships & FKs

Do you display FKs in your LDM / PDM model?
1. Never – They clutter
2. 
3. 
4. 
5. Always – They show meta data

To Party or Not to Party

• Party, Party Role, Party Type, Party Category
• Migration of Party ID
• Use of Subtype “Owned Keys”
• Universality of Party
To Party or Not to Party

- Party
  - PERSON
  - ORGANIZATION
  - PARTY ROLE
  - CUSTOMER
  - VENDOR

- Non-Party
  - DISTRIBUTOR
  - EMPLOYEE
  - VENDOR
  - CUSTOMER

Vote: To Party or Not to Party

Which is the best approach to PARTY in an LDM / PDM?

1. I always use this structure, where applicable.
2.
3.
4.
5. It’s too academic, too theoretical and it just won’t work
Universal Data Models

- Some Data Models are models or subsets of models that an organization can purchase.
- They are sometimes prepared for a specific industry
- May be very generic, may be both.
- May be proprietary or open source

Universal Models

- Silverston UDMs
- ISDMs
- Patterns
- Proprietary Models
- Wiki Models
Vote: Universal Data Models

What do you think about Generic Data Models or Patterns?

1. I love them and won’t ever start a project from a blank page again.
2. 
3. 
4. 
5. They are so high level/cost so much/take too long to understand that I don’t see the value in purchasing them.

SOAs, Ontologies, ESBs, New TLAs and Shoe Strings

• New approaches, methods, techniques
• Often large investments
• Often large returns
• Often large risks
• May be planned with or without data architectures… and architects
SOAs, Ontologies, ESBs, New TLAs and Shoe Strings

- Organizational Maturity
- Methodological Maturity
- Process Maturity
- Professional Maturity
- Managerial Maturity

Vote: New Approaches

How will these new approaches affect your DM goals?
1. More Work / Budget / Focus
2.
3.
4.
5. Less Work / Budget / Focus
Who Gets to Update the Model?

- Many marketing pitches portray a team of developers, modelers, DBAs, etc., happily working on the model
- Is it “More hands make for less work” or “Too many cooks spoil the soup”?

Who Gets to Update the Model

1. Only seasoned Data Modeling Professionals
2. 
3. 
4. 
5. Everybody – we’re all professionals here
Who calls the shots in a Logical Data Model?

- Many stakeholders in a LDM
- Not everyone shares the same understanding about the purpose of the LDM
- DM tools can greatly influence the decision
- DA’s can report through a variety of departments

The FINAL say on LDM / PDM decisions belongs to:

1. The Data Architect/Modeler
2. The Project Manager
3. The User…Customer
4. The DBA
5. The CIO/CEO
Responsibility

• Current Project
• Current Boss
• Enterprise
• Shareholders
• Family
• ???

Responsibility

• If Cost, Benefit, and Risk is our mantra…
• ….whose Costs, Benefits, and Risks do we use?
Vote: To Whom Do We Answer?

1. My Current Boss, on my Current Project, on this day…
2.
3.
4.
5. Everybody (The Public Good)

DA – Lawyer or Architect?

• Several new legislative actions focus on data quality, integrity, reliability, availability…
• Sounds like good DA work, right?
Examples

• Sarbanes-Oxley (SOX): Financial Accountability & Reliability
• Anti-Spam Legislation: Data integrity, Reliability
• Privacy: Quality, Reliability, Accountability…

Vote: DA Role in Compliance

1. Finally, a reason for the Execs to support us…
2.
3.
4.
5. Hey! I don’t even have time write good definitions…
What Are Your Issues?

Is that with one “L” or Two?

How is it spelled?
1. Modeling/Modeler
2. Modelling/Modeller
3. It depends on my mood…
What did we learn?

Near-religious discussions and debates

- People rarely change their minds based on this discussion.
- The best way to convince someone is with measurable facts.
- There is no one “right” solution for all situations.
- The most successful discussions are ones where both sides learn something new about the other viewpoint.
Contentious Issues & Collaboration

…and that we have a long way to go….

If we can’t collaborate well with other DAs, we don’t stand a chance of collaborating well with Project Managers, DBAs, Developers, Testers – or users.
Recommended Books

Thank you!
I'd appreciate feedback of any type about this presentation.
karenlopez@infoadvisors.com
Karen López, ISP

- Karen has 20+ years of data an information architecture experience on large, multi-project programs.
- She is a frequent speaker on data modeling, data-driven methodologies and pattern data models.
- She is a Microsoft MVP, SQL Server
- She wants you to love your data.

Karen López, ISP

- @datachick
- facebook.com/lopezk
- linkedin.com/in/karenlopez

InfoAdvisors

11066 Sheppard Ave East
Toronto, ON  CANADA

karenlopez@infoadvisors.com
http://www.infoadvisors.com

3000+ IRM participants
subscribed to several discussion groups