Data Modeling for the Business Analyst

- Steve Strohl

© 2012, Information Control Corporation
Steve Strohl is a Senior Business Intelligence Architect with over 25 years of IT database experience and 15 years of BI experience. He has been with ICC for the past 7 years. He started his career at Battelle where he worked on defense systems overseas, manufacturing systems and environmental systems. He worked as a lead architect on the Exxon Oil Spill damage assessment database and subsequently lived in Alaska for the next 7 years. Steve worked for an environmental firm where he was able to marry his IT skills with actual field work in Prince William Sound.

“Alaska is where I learned how to ask the question Why instead of just What”.

Steve's work as an adjunct professor at the University of Anchorage also taught him the importance of having a complete view of a problem in order to design an appropriate solution.

Steve returned to Ohio where he began his consulting work in the Business Intelligence field specifically in the Information Delivery field where he designed numerous solutions using Business Objects, Microsoft Analysis Services, Cognos and Microstrategy.

His work in the information delivery field led him very quickly to realize that data quality was a major problem in virtually all of the systems he had experienced.

This knowledge has resulted in his move towards data governance, master data management and the key role that solid business analysis plays in a robust IT solution.

He is currently consulting with Nationwide as the NF program architect for a large master data management initiative and is the Master Data Management Practice Lead @ ICC.

sstrohl@icct.com
It’s a B.A. “Thing” ...

What does the business want?

Have I captured “all” of the business requirements (even those they don’t know about) ?

It’s the Question that Drives Us ...

Have I found all of the hidden complexities ?

How can I represent or relate those needs effectively to my developers and testers ?

Is today garbage day ?
Introduction

**The Goal**
- Establish a process that aids in the discovery process of requirements gathering
- Interface more effectively with developers and tester
- Create deliverables to use throughout the life cycle of the project

**The Steps**
- Start with a Business Problem
- Derive the Requirements
- Derive the Business Rules
- Generate a model that supports the requirements and rules

**The Outcome**
- Better understanding of the business process
- Better interaction with developers and testers
- Better understanding and control over the project
- **BONUS:** Better prepared for CBAP certification.
The B.A.D. - Business Analyst’s Dilemma

How do I ...

✓ Work with client to *discover* the true requirements
✓ Derive the rules that satisfy those requirements
✓ Vet my requirements and rules with the SME’s
✓ Work with the developers to come up with the right solution the first time
✓ Work with the testers to accurately test the solution
✓ Provide documentation that can be used by all parties throughout the project.
Model It! In the end...it’s all about...
The Modeling Methodology

- High Level (Conceptual) Model (HDM)
  - Business Need

- Logical Model (LDM)
  - Business Solution

- Physical Data Model (PDM)
  - Technical Solution

Participants:
- Business Analyst
- Subject Matter Experts
- Architect
- Developer
The Modeling Methodology – The Conceptual Model

High Level (Conceptual) Model (HDM)
(Business Need)

Logical Model (LDM)
Business Solution

Physical Data Model (PDM)
Technical Solution

Business Analyst
Subject Matter Experts

Business Analyst
Architect

Architect
Developer

We lead. You succeed.
# Data Modeling Tool Box

## Entity
- Abstract object (Noun) that describes a unique business meaning
- Contains Title and Definition (optional) of that meaning
- Can be any shape, but a box is standard.

## Relationship Line
- A line that indicates some type of relationship between two entities

## Verb
- A description of the type of relationship between two entities.

## Action Verb
- A description of the type of relationship between two entities.

## Cardinality Indicators
- An indicator of the number of “instances” the relationship has
- 3 Types of cardinality
  - 0: Indicated by a line with a circle
  - 1: Indicated by a line with a smaller perpendicular line
  - M: Indicated by a line with crows feet

## Sub Class
- A symbol that indicates that two or more entities are grouped.
High Level (Conceptual) Model – How it Works

Start with the Business Requirement

Need to track customer orders

Turn a business requirement into a model ...

A customer may place one or more orders.
An order must be placed by only one customer.

Tasks to Complete ...

1. Identify your entities
2. Identify your relationship lines
3. Identify the verb phrases
4. Identify the cardinality
5. Identify any super type/sub type groupings
High Level Model – Define Entities

Turn a business rule into a model …

A customer may place one or more orders.
An order must be placed by only one customer.

Identify the Entities (Find the nouns in the phrase)
A customer may place one or more orders.
An order must be placed by only one customer.

Identify the Entities (Find the nouns in the phrase)
High Level Model – Establish Relationship

Turn a business rule into a model ...

A customer may place one or more orders.
An order must be placed by only one customer.
Turn a business rule into a model ...

A customer may place one or more orders. 
An order must be placed by only one customer.
High Level Model – Define Verbs

Turn a business rule into a model ...

A customer may place one or more orders.
An order must be placed by only one customer.

Identify the Verb Phrase for both directions
Turn a business rule into a model ...

A customer may place one or more orders. An order must be placed by only one customer.

Identify the Verb Phrase from left to right
High Level Model – Define Verbs

Turn a business rule into a model ...

* A customer may place one or more orders.
* An order must be placed by only one customer.

Identify the Verb Phrase from right to left

Customer — places — Order
A customer may place one or more orders.
An order must be placed by only one customer.

Identify the Verb Phrase from right to left
High Level Model – Define Verbs

Turn a business rule into a model ...

A customer may place one or more orders.
An order must be placed by only one customer.

Identify the Verb Phrase for both directions
A customer *may place* one or more orders.
An order must be placed by only one customer.
Turn a business rule into a model ...

A customer *may place one or more orders.*
An order must be placed by only one customer.

**Identify the cardinality from left to right**
High Level Model – Define Cardinality

Turn a business rule into a model ...

A customer may place one or more orders.
An order **must** be placed by **only one** customer.

*Identify the cardinality from right to left*
High Level Model – Define Cardinality

Turn a business rule into a model ...

A customer may place one or more orders.
An order **must** be placed by **only one** customer.

**Identify the cardinality from right to left**
High Level Model – Define Cardinality

Turn a business rule into a model ...

A customer may place one or more orders.
An order must be placed by only one customer.

Identify the cardinality from right to left
How are we doing so far
1st Quiz: Customer and Address

Establish the cardinality ...
Define the business requirement ...

Identify the cardinality between these two entities
What is the business rule ...

*A customer may have more than one address.*

---

**Identify the cardinality between these two entities**
What is the business rule ...

A customer may have more than one address.
An address can be had by more than one customer.

Is this true?

Identify the cardinality between these two entities
2nd Quiz: More Abstract

Turn the phrase into a model ...

*A dog may eat several treats*

*If he does eat any treats he won’t share.*
A dog may eat several treats

If he does eat any treats he won’t share.
High Level Model - Super Type/Sub Types

A Customer is a person or organization ...

A sub type/super type is where you have a natural grouping of two or more entities
Some attributes are shared by all and some attributes are specific to the child member
A Customer is a person or organization ...

_A sub type/super type is where you have a natural grouping of two or more entities_
A Customer is a person or organization ...

A sub type/super type is where you have a natural grouping of two or more entities. A grouping of objects that share attributes, but also have different attributes.

- **Super Type**: Product
- **Sub Types**: Amazon Product, Store Product, Ebay Product

**Non Exclusive**
Can be one or **more**. *(No “X” in the symbol)*
Customer

Address

is

Gold Level

Silver_Level

Bronze Level

Order

Product

had by

has

places

placed by

contained by

contains

is
High Level Conceptual Model – How It’s Useful

How is this useful to me as a Business Analyst...

✓ Identify potentially undiscovered requirements

✓ Implement rules to support requirements

✓ Clear understanding of the relationships of each entity

✓ A working model of the business problem to give to the architect

✓ An effective way to “prove” to the business that you understand the requirements

✓ It will help you to ask some really great questions !!!
The Modeling Methodology – The Logical Model

High Level (Conceptual) Model (HDM) (Business Need)

Logical Model (LDM) Business Solution

Physical Data Model (PDM) Technical Solution

Business Analyst
Subject Matter Experts

Business Analyst
Architect

Developer

Architect

We lead. You succeed.
Next Step: Filling out the Entities

- **Attributes**
  - What are the attributes for that entity

- **Keys**
  - What uniquely identifies a record in that entity

- **Resolve M:M**
  - Break out many to many entities
### A Short Glossary of Terms

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Attribute</strong></td>
<td>• An is a descriptor that uniquely identifies a record in the entity</td>
</tr>
<tr>
<td><strong>Primary Key</strong></td>
<td>• An attribute or set of attributes that uniquely identify a record in an entity</td>
</tr>
<tr>
<td></td>
<td>• There can be only one primary key, but more than one unique key</td>
</tr>
<tr>
<td></td>
<td>• This value can never be empty or null</td>
</tr>
<tr>
<td><strong>Alternate or Natural Key</strong></td>
<td>• An attribute or set of attributes that has business meaning and uniquely identifies a record</td>
</tr>
<tr>
<td></td>
<td>• This can be the primary key to the entity if no surrogate key has been defined</td>
</tr>
<tr>
<td></td>
<td>• This value can never be empty or null</td>
</tr>
<tr>
<td><strong>Surrogate Key</strong></td>
<td>• An attribute or set of attributes that has no business meaning.</td>
</tr>
<tr>
<td></td>
<td>• It is a sequential number that uniquely identifies a record.</td>
</tr>
<tr>
<td></td>
<td>• This is usually the primary key, the alternate key will be a unique key</td>
</tr>
<tr>
<td></td>
<td>• This value can never be null and is usually found in the physical model only</td>
</tr>
<tr>
<td><strong>Foreign Key</strong></td>
<td>• An attribute or set of attributes that is the unique key from another entity. These two entities are now linked.</td>
</tr>
</tbody>
</table>
Customer – Address Entity Walk Thru

CUSTOMER

IDENTIFY CUSTOMER AND ADDRESS ATTRIBUTES

ESTABLISH NATURAL KEYS (Always do Natural Keys first)

ESTABLISH SURROGATE KEY

IDENTIFY WHICH KEY IS THE PRIMARY KEY

IDENTIFY FOREIGN KEYS

RESOLVE RELATIONSHIP

ADDRESS

Customer

Attributes
Alternate Key
Surrogate Key

Address

Attributes
Alternate Key
Surrogate Key

Relationships

1:M
M:M
Establish Customer Attributes

**Attributes**

• A collection of items that describes the entity
Establish Customer Attributes

**Attributes**

- A collection of items that describes the entity

<table>
<thead>
<tr>
<th>CUSTOMER</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSN</td>
</tr>
<tr>
<td>Prefix</td>
</tr>
<tr>
<td>Last Name</td>
</tr>
<tr>
<td>First Name</td>
</tr>
<tr>
<td>Middle Initial</td>
</tr>
<tr>
<td>Suffix</td>
</tr>
<tr>
<td>DOB</td>
</tr>
</tbody>
</table>

**Customer**

- Attributes
- Alternate Key
- Surrogate Key

**Address**

- Attributes
- Alternate Key
- Surrogate Key

**Relationships**

- 1:M
- M:M
**Establish Customer Unique Key**

### Alternate or Natural Key(s)

- An attribute or attributes that uniquely identifies a record in the entity.
- These attributes must have meaning to the record.
- AK means “Alternate Key” also called a “Natural Key”
- This attribute can never be empty or null

### Questions:

*What is a good natural key or keys for this entity?*
Establish Customer Unique Key

**Alternate or Natural Key(s)**

- An attribute or attributes that uniquely identifies a record in the entity.

- These attributes must have meaning to the record.

- AK means “Alternate Key” also called a “Natural Key”

- This attribute can never be empty or null
Establish Customer Surrogate Key

**Surrogate Key(s)**
- An “non intelligent” system generated number
- This attribute is generated for each record and is unique
- If a surrogate key is used it is the primary key.
  - If it is not used the natural key becomes the primary key
- A surrogate key is used for performance reasons.
- This key value can never be empty or null

**Attributes**
- SSN
- Prefix
- Last Name
- First Name
- Middle Initial
- Suffix
- DOB

**Alternate Key**
- An "non intelligent" system generated number
- This attribute is generated for each record and is unique
- If a surrogate key is used it is the primary key.
- If it is not used the natural key becomes the primary key
- A surrogate key is used for performance reasons.
- This key value can never be empty or null

**Address**
- Attributes
- Alternate Key
- Surrogate Key

**Relationships**
- 1:M
- M:M

**What will the surrogate key be?**

**A non intelligent attribute is an attribute that has no business meaning to the entity.**
Establish Customer Surrogate Key

<table>
<thead>
<tr>
<th>CUSTOMER</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PK</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Customer ID</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AK</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prefix</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Last Name</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First Name</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle Initial</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suffix</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOB</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Surrogate Key(s)**

- An “non intelligent” system generated number
- This attribute is generated for each record and is unique
- If a surrogate key is used it is the primary key.
- If it is not used the natural key becomes the primary key
- A surrogate key is used for performance reasons.
- This key value can never be empty or null

**A non intelligent attribute is an attribute that has no business meaning to the entity**

If you have a surrogate key do you still need to identify the natural key? Why?
Establish Address Attributes

**CUSTOMER**

- **PK**
  - Customer ID
- **AK**
  - SSN
  - Prefix
  - Last Name
  - First Name
  - Middle Initial
  - Suffix
  - DOB

**ADDRESS**

**Customer**
- Attributes
- Alternate Key
- Surrogate Key

**Address**
- Attributes
- Alternate Key
- Surrogate Key

**Relationships**
- 1:M
- M:M
## Establish Address Attributes

<table>
<thead>
<tr>
<th>CUSTOMER</th>
<th>ADDRESS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PK</strong></td>
<td></td>
</tr>
<tr>
<td>Customer ID</td>
<td>Address Line 1</td>
</tr>
<tr>
<td><strong>AK</strong></td>
<td>Address Line 2</td>
</tr>
<tr>
<td>SSN</td>
<td>City</td>
</tr>
<tr>
<td>Prefix</td>
<td>State</td>
</tr>
<tr>
<td>Last Name</td>
<td>Zip</td>
</tr>
<tr>
<td>First Name</td>
<td>Zip + 4</td>
</tr>
<tr>
<td>Middle Initial</td>
<td></td>
</tr>
<tr>
<td>Suffix</td>
<td></td>
</tr>
<tr>
<td>DOB</td>
<td></td>
</tr>
</tbody>
</table>

### Attributes
- PK
- AK
- SSN
- Customer ID
- Address Line 1
- Address Line 2
- City
- State
- Zip
- Zip + 4

### Relationships
- 1:M
- M:M
Establish Address Natural Key

What are good natural keys for the ADDRESS entity?
Establish Address Natural Key

<table>
<thead>
<tr>
<th>CUSTOMER</th>
<th>ADDRESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>PK</td>
<td>AK</td>
</tr>
<tr>
<td>Customer ID</td>
<td>Address Line 1</td>
</tr>
<tr>
<td>SSN</td>
<td>AK</td>
</tr>
<tr>
<td>Prefix</td>
<td>Address Line 2</td>
</tr>
<tr>
<td>Last Name</td>
<td>AK</td>
</tr>
<tr>
<td>First Name</td>
<td>City</td>
</tr>
<tr>
<td>Middle Initial</td>
<td>State</td>
</tr>
<tr>
<td>Suffix</td>
<td>AK</td>
</tr>
<tr>
<td>DOB</td>
<td>AK</td>
</tr>
<tr>
<td></td>
<td>AK</td>
</tr>
<tr>
<td></td>
<td>AK</td>
</tr>
</tbody>
</table>

Does this make sense?
Establish Address Surrogate Key

What is the surrogate key for the ADDRESS entity?
Establish Address Surrogate Key

<table>
<thead>
<tr>
<th>CUSTOMER</th>
<th>ADDRESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>PK</td>
<td>PK</td>
</tr>
<tr>
<td>AK</td>
<td>AK</td>
</tr>
<tr>
<td>Customer ID</td>
<td>Address ID</td>
</tr>
<tr>
<td>SSN</td>
<td>Address Line 1</td>
</tr>
<tr>
<td>Prefix</td>
<td>Address Line 2</td>
</tr>
<tr>
<td>Last Name</td>
<td>City</td>
</tr>
<tr>
<td>First Name</td>
<td>State</td>
</tr>
<tr>
<td>Middle Initial</td>
<td>Zip</td>
</tr>
<tr>
<td>Suffix</td>
<td>Zip + 4</td>
</tr>
<tr>
<td>DOB</td>
<td></td>
</tr>
</tbody>
</table>

**Attributes**
- Customer: PK, AK
- Address: PK, AK

**Alternate Key**
- Customer: AK
- Address: AK

**Surrogate Key**
- Customer: PK
- Address: PK

**Customer Relationships**
- 1:M
- M:M
Establish Foreign Key (1:1 or 1:M)

What key from the CUSTOMER entity will be the foreign key in the ADDRESS entity?
Establish Foreign Key (1:1 or 1:M)

**CUSTOMER**
- PK: Customer ID
- AK: SSN
  - Prefix
  - Last Name
  - First Name
  - Middle Initial
  - Suffix
  - DOB

**ADDRESS**
- PK: Address ID
- FK: Customer ID
- AK: Address Line 1
- AK: Address Line 2
- AK: City
- AK: State
- AK: Zip
- AK: Zip + 4

**Foreign Key**
A foreign key is a primary or unique key from another table
Establish Foreign Key (M:M)

**CUSTOMER**
- PK: Customer ID
- AK: SSN
  - Prefix
  - Last Name
  - First Name
  - Middle Initial
  - Suffix
  - DOB

**ADDRESS**
- PK: Address ID
- AK: Address Line 1
- AK: Address Line 2
- AK: City
- AK: State
- AK: Zip
- AK: Zip + 4

**Customer**
- Attributes
- Alternate Key
- Surrogate Key

**Address**
- Attributes
- Alternate Key
- Surrogate Key

**Relationships**
- 1:M
- M:M
Establish Foreign Key (M:M)

**CUSTOMER**
- **PK**: Customer ID
- **AK**: SSN
  - Prefix
  - Last Name
  - First Name
  - Middle Initial
  - Suffix
  - DOB

**ADDRESS**
- **PK**: Address ID
- **AK**: Address Line 1
- **AK**: Address Line 2
- **AK**: City
- **AK**: State
- **AK**: Zip
- **AK**: Zip + 4

**CUSTOMER_ADDRESS**

**Relationships**
- 1:M
- M:M

**Attributes**
- Yes
  - Customer
    - Attributes
    - Alternate Key
    - Surrogate Key
  - Address
    - Attributes
    - Alternate Key
    - Surrogate Key

© 2012, Information Control Corporation
Establish Foreign Key (M:M)

**Customer**
- Attributes
- Alternate Key
- Surrogate Key

**Address**
- Attributes
- Alternate Key
- Surrogate Key

**Relationships**
- 1:M
- M:M

Why did we add the “Address Type” attribute?
The Modeling Methodology – Hand Off to Architect

Hand Off to the Architect. Architect will save valuable time in development and testing.
Logical Model – How It’s Useful

How is this useful to me as a Business Analyst...

✓ Thorough knowledge of the business entities and keys and attributes

✓ Identification of potential gaps in the application where uniqueness is not enforced

✓ Identification of “Bridge” entity and attributes between M:M entities

✓ A blue print that the architect, modeler or DBA can use to build a physical model

✓ A working document that testers can use to build effective test cases.

✓ A common blue print that can be used for discussion with the business and IT
Questions

Address

had by

has

Customer

places

placed by

Order

contains

contained by

Product

Gold Level

Silver Level

Bronze Level

is

is

We lead. You succeed.