

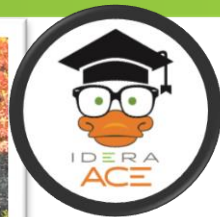
Data Integrity Demystified

Deborah Melkin
she\her
March 10, 2021





About Me



- 20 years as a DBA
 - Mainly work with SQL Server
 - Mainly work with OLTP but have worked with some data marts.
 - NESQL Board Member
 - SQL Saturday\ User Group Speaker
 - IDERA ACE Class of 2020
 - Speaker Idol Winner 2019
 - Microsoft MVP – Data Platform
- Random facts:**
- I'm the alto section leader in my choir.
 - I go to bluegrass jams regularly.
 - I've been learning guitar and now mandolin.
 - I am a bit of a musical theater geek.

Agenda

- Setting the Foundation
- Designing for the Data
- Data Integrity & Performance



Let's Define "Integrity"



“Integrity”

integrity noun



Save Word

in·teg·ri·ty | \ in- 'te-grə-tē  \

Definition of *integrity*

- 1 : firm adherence to a code of especially moral or artistic values : INCORRUPTIBILITY
- 2 : an unimpaired condition : SOUNDNESS
- 3 : the quality or state of being complete or undivided : COMPLETENESS

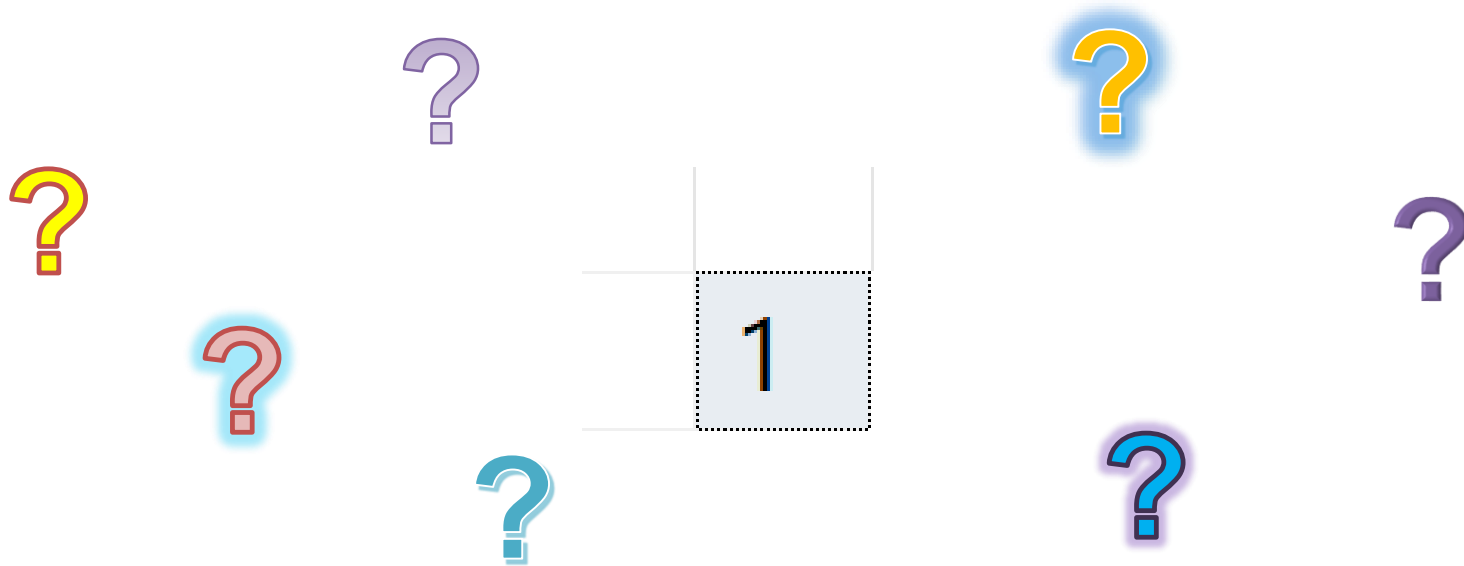
<https://www.merriam-webster.com/dictionary/integrity>

“Data Integrity”

...the maintenance of, and the assurance of, the accuracy and consistency of data over its entire life-cycle, and is a critical aspect to the design, implementation and usage of any system which stores, processes, or retrieves data....

https://en.wikipedia.org/wiki/Data_integrity

What Does It Mean???



Why Does This Matter?

- “Garbage In – Garbage Out”
- Trust of data
- Not independent of other data
- Accessed by multiple sources
- Solve problems before they happen

Setting the Foundation...



Data Storage & Data Engines

- Relational Database
- NoSQL Database
- Data Lake
- Excel
- Text File
- etc....




Naming Conventions

- Capitalization and punctuation
- Character limits
- Singular or Plural table name
- TLA (three letter aliases) or other prefixes
- Schemas

Naming Convention Examples


CamelCase

 CamelCaseId


 CamelCaseName


lowercase

 id

 alllowercase

underscore_between_words

 underscore_ID

 underscore_column_name


Supercalifragilisticexpialidocious_is_thirty_four_characters_long


long_nonsense_word


Gadget


Gadgets


History_of_Names


 NamesID

 original_convention

 changewith2014release

 ChangeWith2016Hotfix5

 Next_Change_In_Convention_2018

 AnythingGoes_2020

Foundations Create Integrity By...

- Defining the structure of the data
- Understanding the type of data at the start
- Looking at names to easily understand what the data represents by
- Creating consistency and a structure to the definitions used in the database

Designing for the Data...



Different Types of Integrity

- Entity Integrity
 - About the table (entity)
- Referential Integrity
 - About the relationships between tables/entities
- Domain Integrity
 - About the column (attribute)

Entity Integrity



Relational Database Design

- Normal Form vs. Star Schema
- Determines how to group the different pieces of data in a logical way
- Provides additional meaning based on the column's "role" on the table

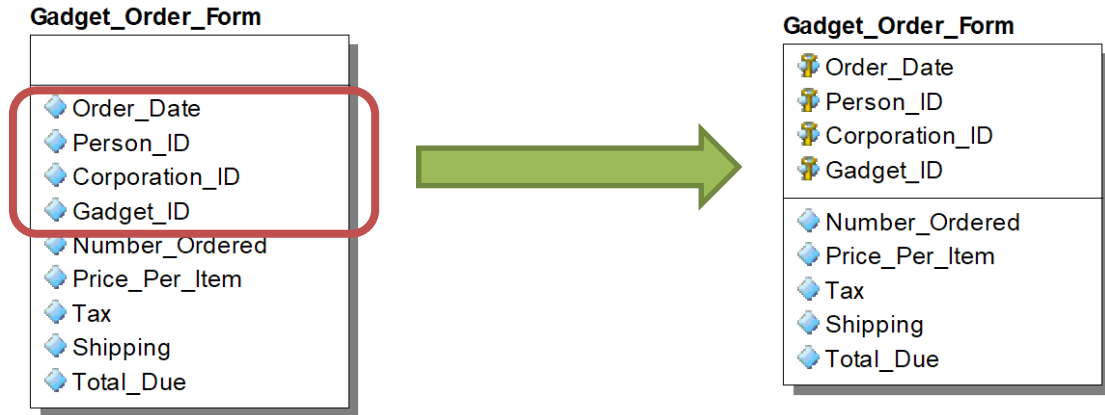
Primary Key Constraint

- Column or combination of columns that define the uniqueness of each record of the table
- All columns must be NOT NULL
- Only 1 primary key per table
- Logical vs Physical

Tip: Identify Primary Key Column(s)

Ask the following question:

What does this table hold using 1 or 2 sentences?



Surrogate Primary Key

surrogate noun, often attributive

sur-ro·gate | \ˈsər-ə-,gāt, ˈsə-rə-, -gət \

Definition of *surrogate* (Entry 2 of 2)

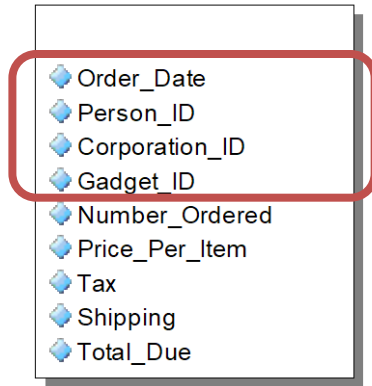
- 1 a : one appointed to act in place of another : DEPUTY
b : a local judicial officer in some states (such as New York) who has jurisdiction over the probate of wills, the settlement of estates, and the appointment and supervision of guardians
- 2 : one that serves as a substitute
- 3 : SURROGATE MOTHER

<https://www.merriam-webster.com/dictionary/surrogate>

Do Surrogate PKs Create Integrity?

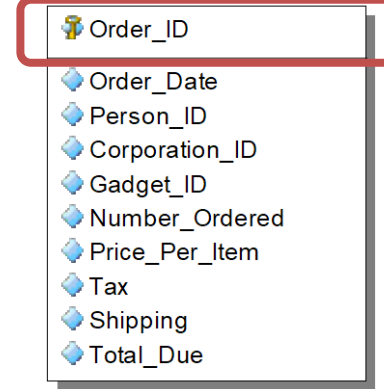
They Don't!

Gadget_Order_Form



◆ Order_Date
◆ Person_ID
◆ Corporation_ID
◆ Gadget_ID
◆ Number_Ordered
◆ Price_Per_Item
◆ Tax
◆ Shipping
◆ Total_Due

Gadget_Order_Form



◆ Order_ID
◆ Order_Date
◆ Person_ID
◆ Corporation_ID
◆ Gadget_ID
◆ Number_Ordered
◆ Price_Per_Item
◆ Tax
◆ Shipping
◆ Total_Due

Do Surrogate PKs Create Integrity?

~~“If you stand for nothing, you will fall for anything.”~~

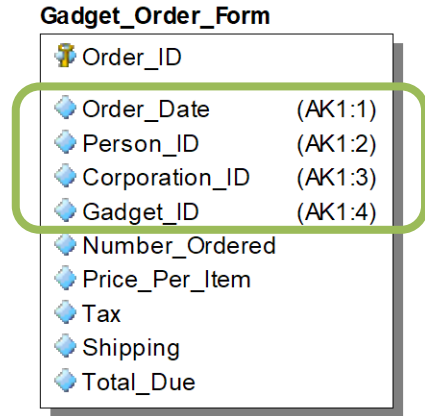
If your surrogate primary key stands for nothing,
your data integrity will “fall.”

Unique Key Constraint

- Column or combinations of columns that create a unique set in addition to the Primary Key.
- Not as strict as Primary Key – allows for NULL columns.
- Can have multiple unique combinations.

Unique Key Constraint (cont'd)

- Difference between a Unique Key and a Unique Index:
 - Use of filtered indexes
 - Use of included columns



The screenshot shows a table named 'Gadget_Order_Form' with the following columns:

Order_ID	
Order_Date	(AK1:1)
Person_ID	(AK1:2)
Corporation_ID	(AK1:3)
Gadget_ID	(AK1:4)
Number_Ordered	
Price_Per_Item	
Tax	
Shipping	
Total_Due	

A green rectangular box highlights the first four columns: Order_ID, Order_Date, Person_ID, and Corporation_ID.

What does Entity Integrity do?

- Defines the relationship between...
 - the columns on the table
 - the tables themselves
- Creates a clear definition of what each record is about

What does Entity Integrity do?

- Creates all of the alternate primary key combinations or the logical primary key in case of a surrogate primary key
- Enforces additional business rules

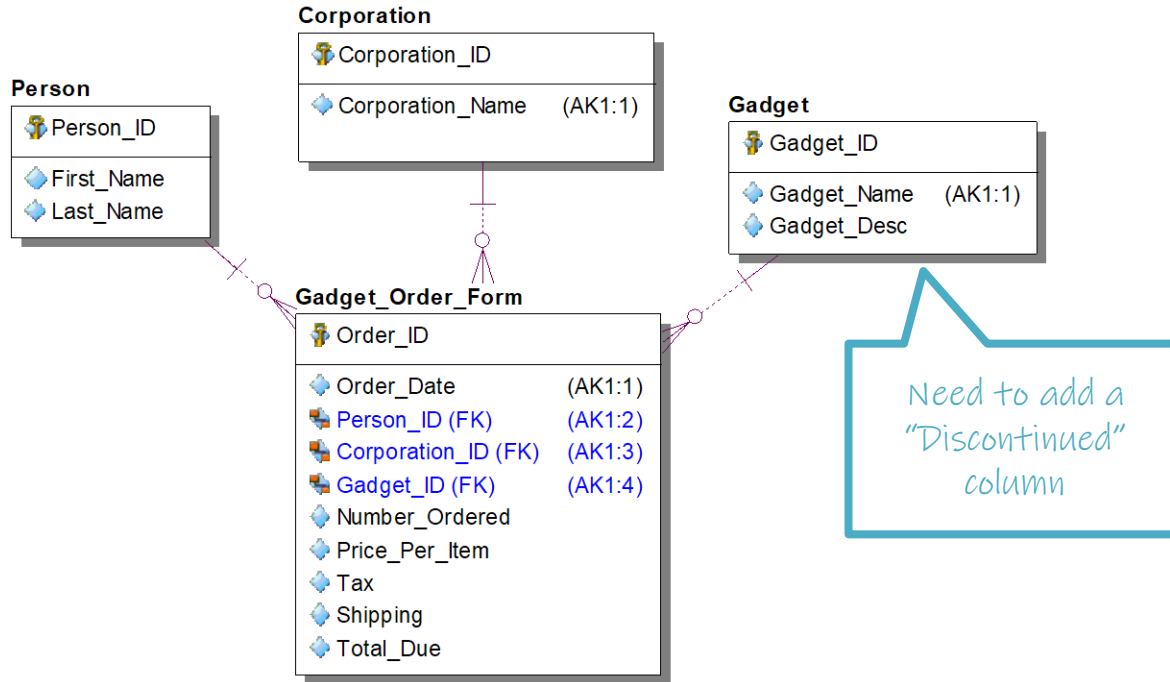
Referential Integrity



Foreign Key Constraint

- Creates relationship between two tables
- Ensures that the value in the field exists in the referenced table
- Can define actions that occur when the referenced column is changed or the row is removed

Foreign Key Example



What does Referential Integrity do?

- Ensures that data that is related to other tables is valid
- Defines rules for handling changes with the referenced column to maintain the data integrity

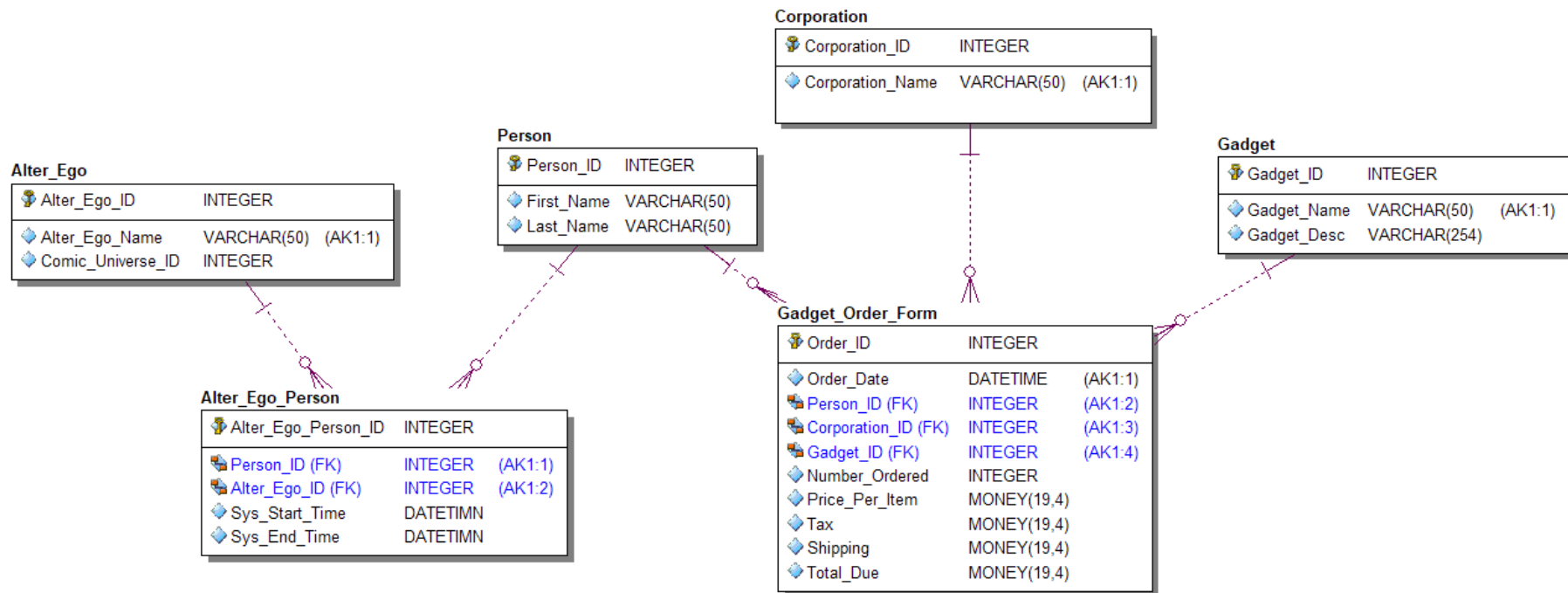
Domain Integrity



Data Types

- Sets the type of data we are dealing with
- Creates constraints on the data goes inside the attribute
- Can guarantee valid data

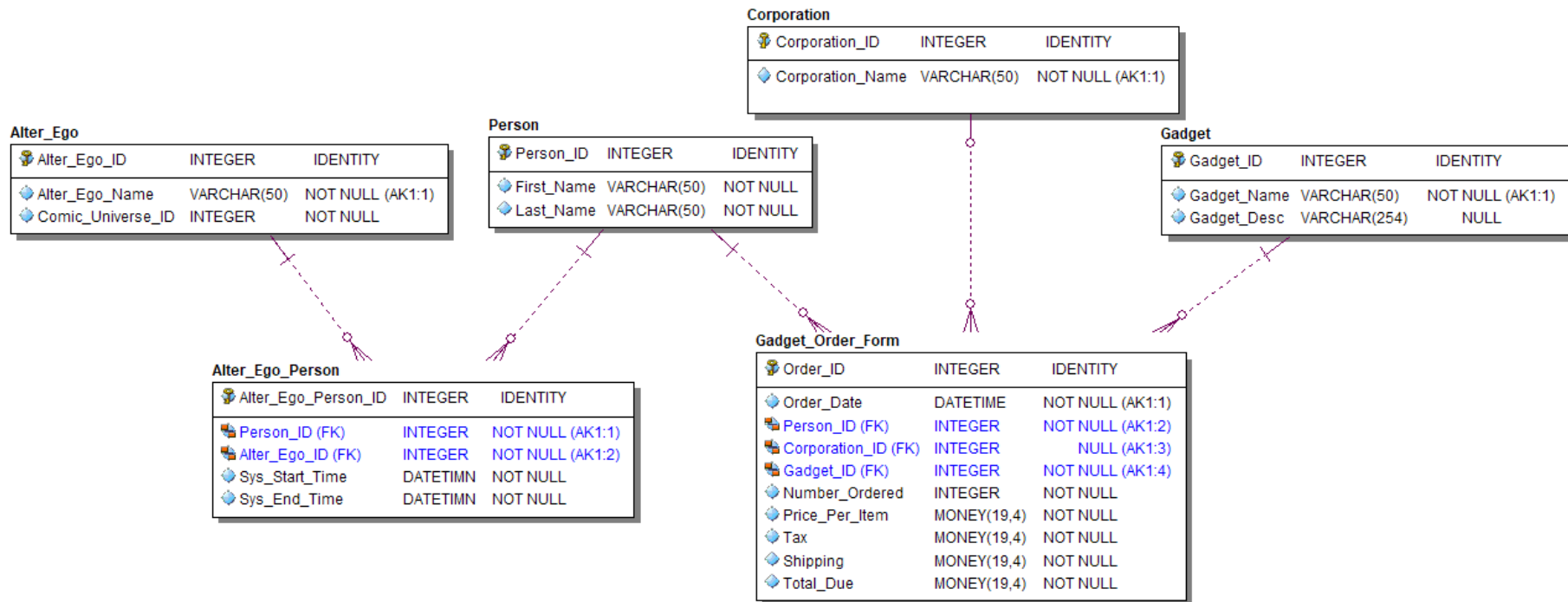
Data Types



NOT NULL

- NULL is the absence of a value
- Sets the requirement as to whether the attribute must have a value
- " (empty string) \neq NULL






NOT NULL



Default Constraint

- Set the value for a column if the value for the column is not specified
- Ignored if the value is specified as NULL
- Applies to Inserts or Foreign Key with SET DEFAULT

Alter_Ego_Person

 Alter_Ego_Person_ID	INTEGER	IDENTITY	
 Person_ID (FK)	INTEGER	NOT NULL	(AK1:1)
 Alter_Ego_ID (FK)	INTEGER	NOT NULL	((1)) (AK1:2)
 Sys_Start_Time	DATETIME	NOT NULL	
 Sys_End_Time	DATETIME	NOT NULL	

Check Constraint

- Creates restrictions on the data that goes in the column
- Ensuring the data in the field matches a given criteria
- Can be created on the column or table level
 - Table level means that multiple columns can be checked at once

What Does Domain Integrity Do?

- Defines the type of data expected for each column
- Ensures the data in the column matches what should be there
- Sets whether the field is required to have a value

What Does Domain Integrity Do?

- Establishes the common known value for the field when the record is first inserted
- Ensures that only the values that we expect to find in the column (or table) are there
- Eliminates a way for bad data to get in our tables

Data Integrity & Performance



Clustered Index: Integrity or Performance

Clustered index defines the order of how the data in the table is physically stored.

Integrity

Performance

Integrity and the Query Optimizer

- NOT NULL
- Primary Key Constraints
- Unique Constraints
- Foreign Key Constraints
- Check Constraints

Data Types Performance

- Database size bloat
- Memory Grant
- Implicit Conversions

Demos!



Parting Thoughts...

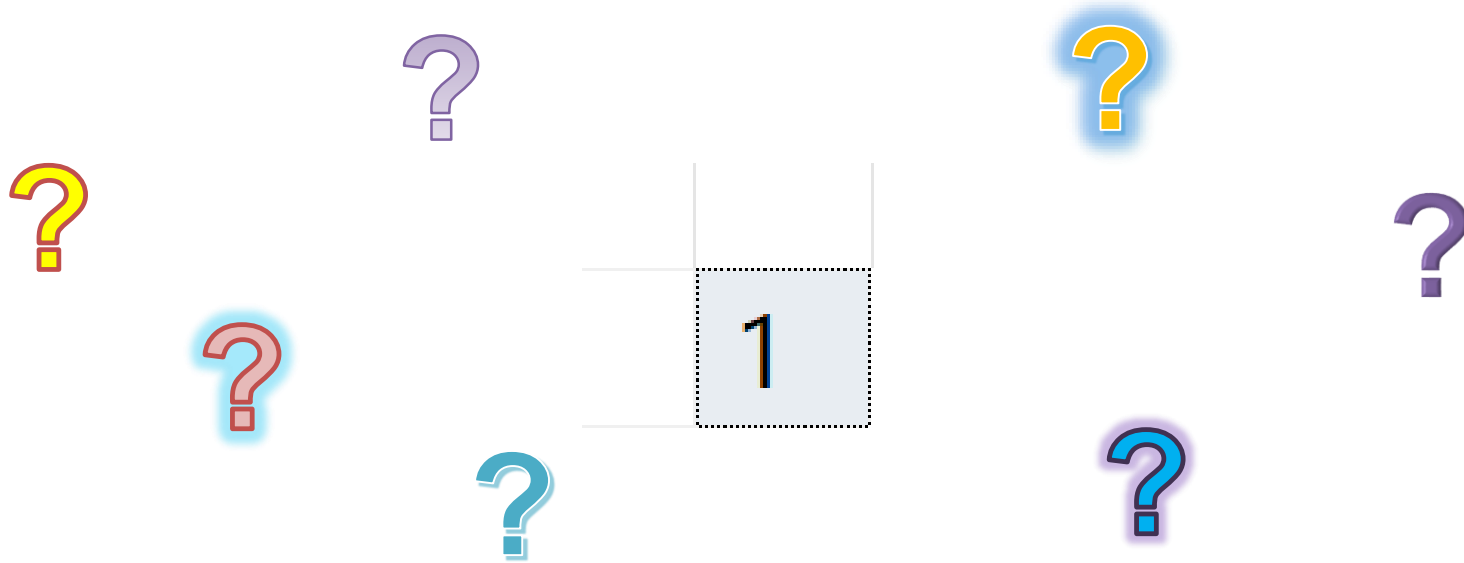


Is Data Integrity 100% Achievable?

No!

- But it is worthwhile to figure out what is needed to create true data integrity to better understand and work with your data.

Remember this?



So What Did It Stand for?

```
1 SELECT P.First_Name, P.Last_Name,  
2       G.Gadget_Name, GOrder.Number_Ordered  
3 FROM Gadget_Order_Form as GOrder  
4     JOIN Gadget as G on GOrder.Gadget_ID = G.Gadget_ID  
5     JOIN Person as P ON GOrder.Person_ID = P.Person_ID  
6 WHERE g.Gadget_ID IN (2,3)  
7
```

100 %

Results Messages

	First_Name	Last_Name	Gadget_Name	Number_Ordered
1	Deborah	Melkin	Magical Bracelets	2
2	Deborah	Melkin	Royal Tiara	1



References

- <https://docs.microsoft.com/en-us/sql/relational-databases/tables/unique-constraints-and-check-constraints?view=sql-server-ver15>
- <https://littlekendra.com/2016/09/08/unique-constraints-vs-unique-indexes/>

References (cont'd)

- <https://debthedba.wordpress.com/2018/10/31/how-i-really-feel-about-surrogate-primary-keys/>
- <https://www.scarydba.com/2015/09/09/yes-foreign-keys-help-performance/>
- <https://www.sqlservercentral.com/articles/trust-or-verify-%E2%80%93-how-fks-can-or-cannot-help-performance>

Email: dgmelkin@gmail.com

Twitter: @dgmelkin

Blog: DebtheDBA.wordpress.com

GitHub: <https://tinyurl.com/y5dpvue2>

Thank you for Coming!

