THE KEY TO KEYS

PRIMARY ALTERNATE IDENTITY UNIQUE CLUSTERED SEQUENCE SURROGATE CLUTTERED FOREIGN MEANINGLESS HASH ROWIDENTIFIER ROWID LOOKUP SEARCH PK FK GUID BUSINESS IDENTIFIER PROPERTY CONSTRAINT INDEX

Karen Lopez @datachick

#TeamData

YES, PLEASE DO TWEET/SHARE TODAY'S EVENT

@DATACHICK #TEAMDATA

KAREN LÓPEZ

Karen has 20+ years of data and information architecture experience on large, multi-project programs.

She is a frequent speaker on data modeling, data-driven methodologies and pattern data models.

She wants you to love your data…

She is loves new tech and gadgets

MVP Microsoft® Most Valuable Professional

@datachick

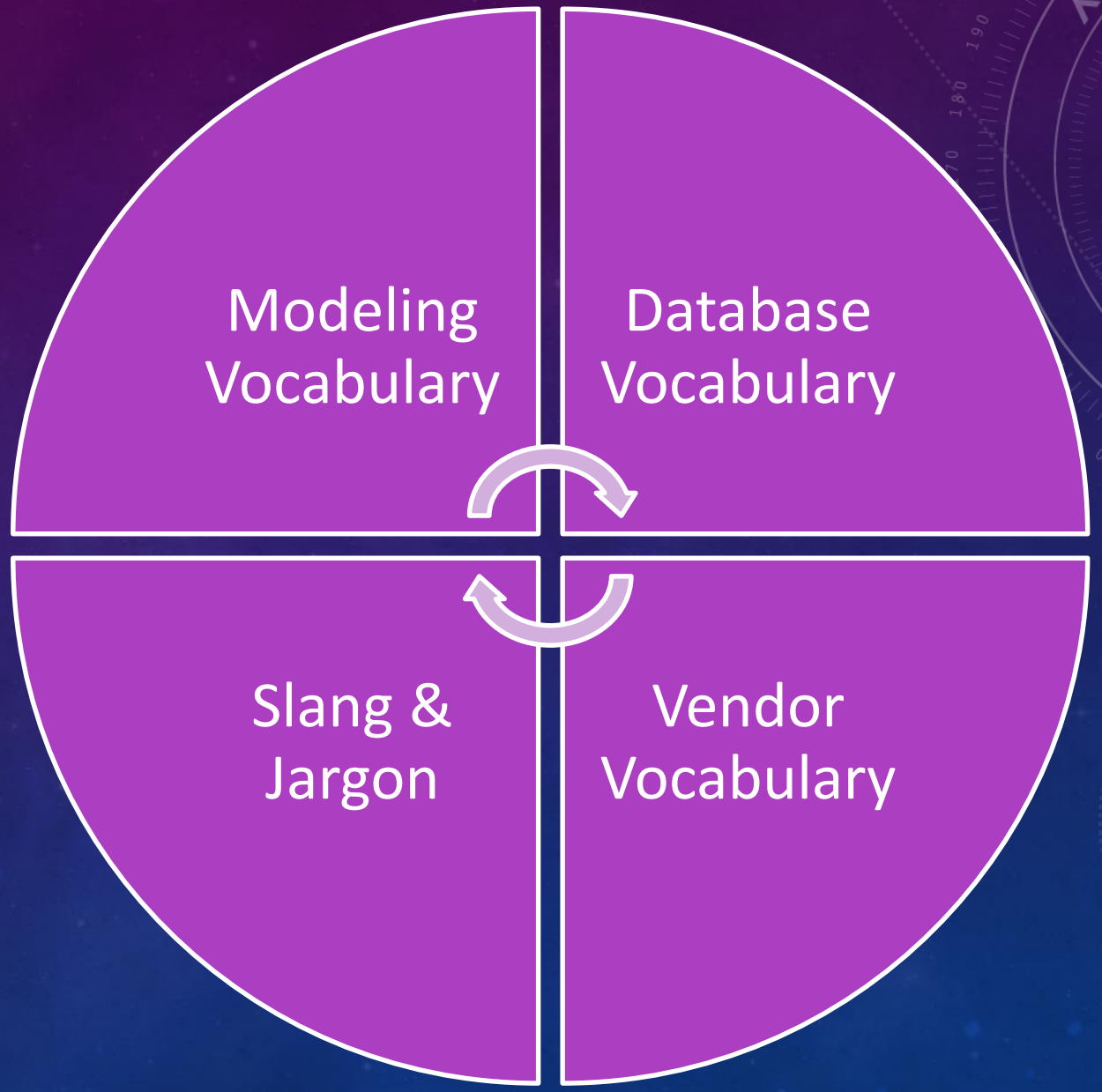
facebook.com/lopezk

linkedin.com/in/karenlopez

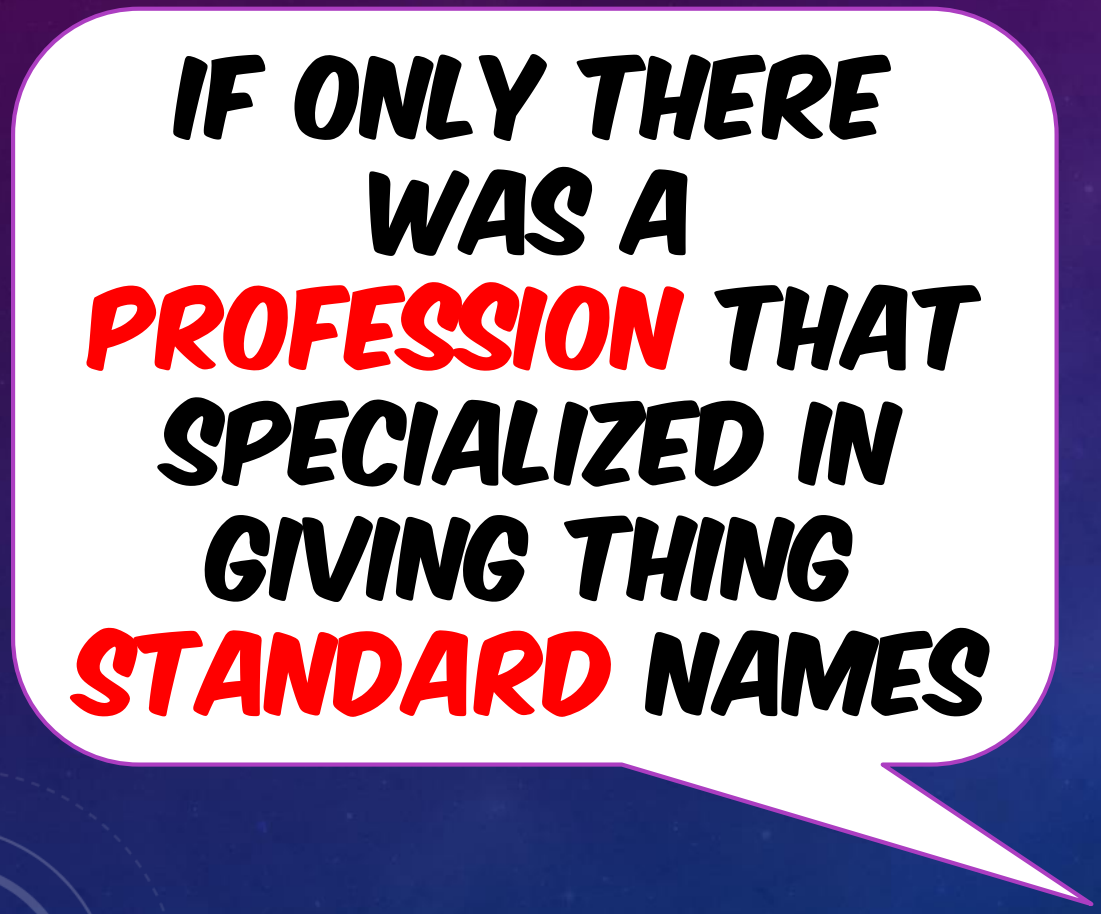
WHO ARE YOU?

SO MANY TERMS AND JARGON IN TALKING ABOUT DATABASE STUFF

WHY THIS TOPIC?



KAREN SAYS: KEYS



1. Keys are Key to performance AND data quality
2. Concepts can have many ways to implement
3. Not all data modeling vendors use the same terms
4. Not all database vendors use are the same

TODAY'S FOCUS

Entities and Tables

Attributes and Columns

Database objects that implement or enforce key-related stuff

ONE MORE THING… KEYS ARE KEY!

Keys (Identifiers) play a role that many modelers & designers forget:	They form part of the semantics of our models.
	Primary keys are used in enforcing constraints on data quality
	Primary keys are not just about performance.

NORMALIZATION

- Is all about the keys, ‘bout the keys, ‘bout the keys…
- Depends on understanding the MEANING of the keys and columns
- Goes all to heck* when you have surrogate keys
- Depends on the make up of the key parts (columns)

3NF

Every Fact is either part of a key or depends upon the key, the whole key, and nothing but the key.

….so help you Codd

Michael J Swart

LINGO, JARGON AND TERMS

WE ARE LOUSY AT STANDARDIZING THESE THINGS

TYPES OF KEY VOCABULARIES

Terms used conceptually

- Primary key
- Alternate key
- Composite key
- Super key
- Candidate key
- Surrogate key
- …

Terms used physically

- Primary key
- Clustered key
- Encryption key
- Partitioning key
- Index
- Identity
- Sequence

PRIMARY

Foreign

Unique

Clustered

Partitioning

GUID

Index

Sequence

Surrogate

Encryption

Super

Alternate

Lookup

Identity

Candidate

Business

Constraint

Logical

Natural

Compound

Numeric

Composite

Cluttered

Duplicate

AT THE BEGINNING

Business Key

Logical Key

Natural Key

CLOSER TO DESIGN…

- Super Key
- Candidate Key
- Primary Key
- Alternate Key
- Foreign Key

PRIMARY KEY CRITERIA

- **Applicable to all instances (Mandatory)**
- **Unique**
- **Stable**
- **Small**

The first two are required by the Relational Model.

3 and 4 required by good practice. But they are not required in data modeling.

Modelers should, however, live in the real world most of the time and observe them

ALTERNATE KEYS

Unique

Mandatory

Might Change

Unique Constraint

Unique Index

MORE TERMS

Cluttered key

Composite key

Compound key

MORE TERMS

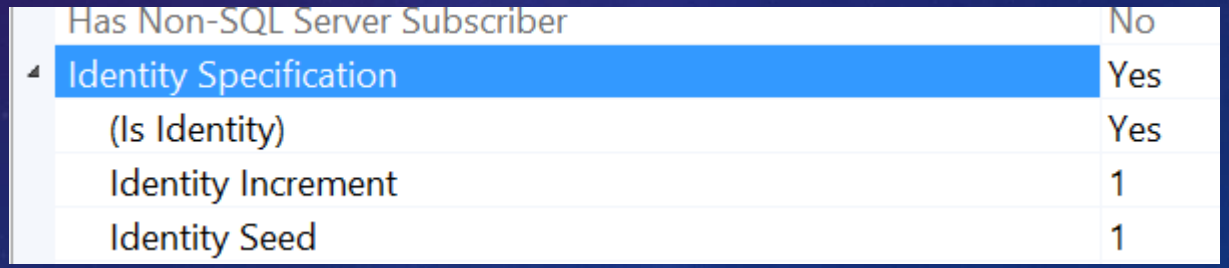
Surrogate Key

Identity

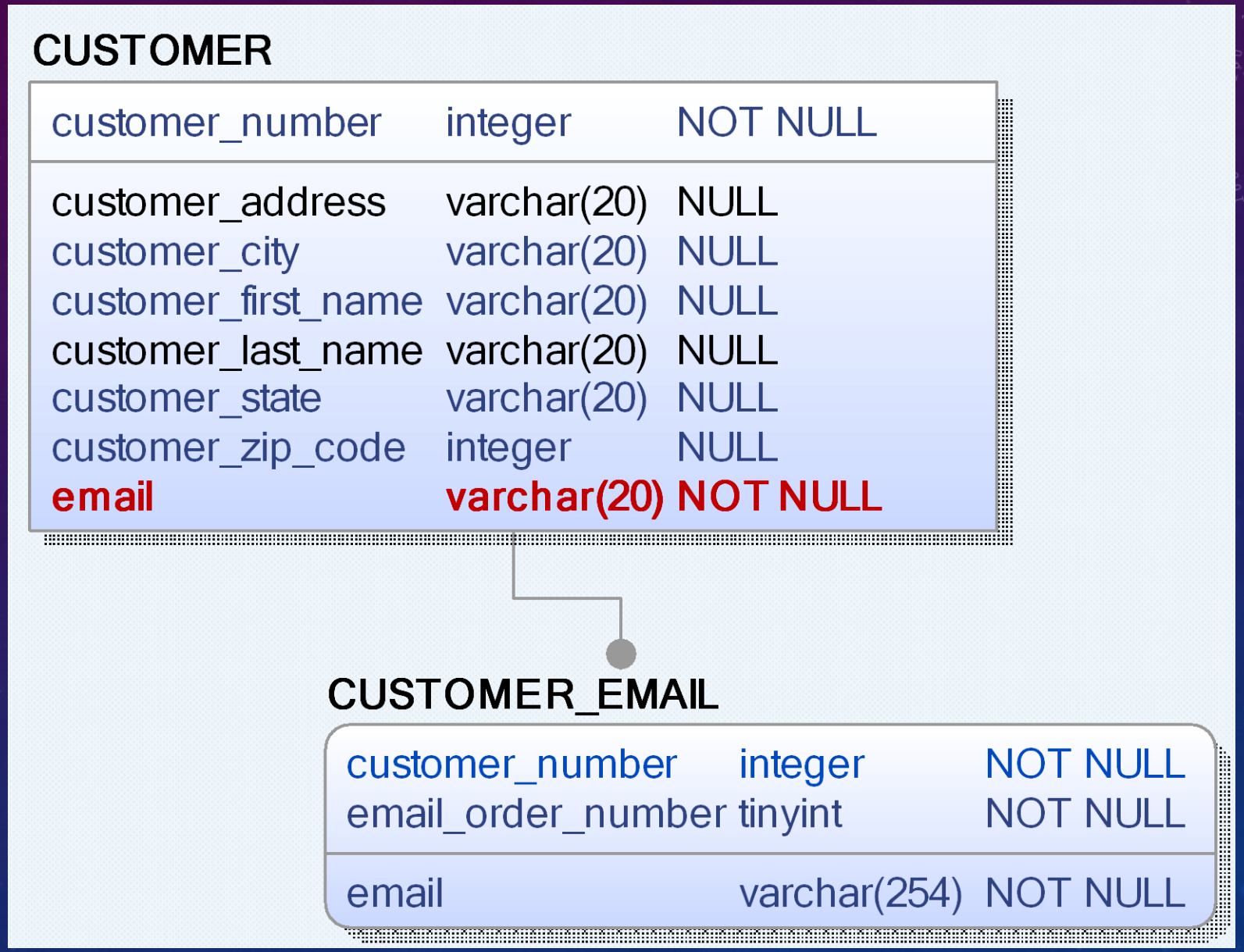
Sequence

GUID

Custom



SURROGATE KEYS


CUSTOMER
customer_number integer NOT NULL
customer_address varchar(20) NULL
customer_city varchar(20) NULL
customer_first_name varchar(20) NULL
customer_last_name varchar(20) NULL
customer_state varchar(20) NULL
customer_zip_code integer NULL
email varchar(20) NOT NULL
CUSTOMER_EMAIL
customer_number integer NOT NULL
email_order_number tinyint NOT NULL
email varchar(254) NOT NULL

IDENTITY/IDENTITY PROPERTY

Issue: How people use them

```
IDENTITY [ (seed , increment) ]
```

2,147,483,647

9,223,372,036,854,775,807

32,767

255

IDENTITIES

Issue: How people monitor them

IDENTITY [(seed , increment)]

HOW DO YOU FIX IT IF IT'S ALREADY GONE OFF?

RESEED

GAPS

Not UNIQUE?

Change DataType

WHAT ABOUT SEQUENCES?

```
CREATE SEQUENCE [schema_name . ] sequence_name
[ AS [ built_in_integer_type | user-defined_integer_type ] ]
[ START WITH <constant> ]
[ INCREMENT BY <constant> ]
[ { MINVALUE [ <constant> ] } | { NOMINVALUE } ]
[ { MAXVALUE
[ <constant> ] } | { NOMAXVALUE } ]
[ CYCLE | { NOCYCLE } ]
[ { CACHE [ <constant> ] } | { NO CACHE } ]
[ ; ]
```

RESTART

```
ALTER SEQUENCE [schema_name. ] sequence_name
[ RESTART [ WITH <constant> ] ]
[ INCREMENT BY <constant> ]
[ { MINVALUE <constant> } | { NO MINVALUE } ] [ { MAXVALUE <constant> }
| { NO MAXVALUE } ]
[ CYCLE | { NO CYCLE } ] [ { CACHE [ <constant> ] } | { NO CACHE } ]
 [ ; ]
```

HOW DO YOU FIX IT IF IT'S ALREADY GONE OFF?

RESTART

GAPS

Not UNIQUE?

Change DataType

HOW DO YOU PREVENT IT?

Proper Design

- Do the math
- SEED Value
- Increment Value
- Use another feature

Monitor

- Max allowable value
- Max current value
- Gaps

Predict* countdown timer

- Team Data Script
- Know your hottest tables
- Prioritize

http://thomaslarock.com/2015/11/sql-server-identity-values-check/

GUIDS

3F2504E0-4F89-41D3-9A0C-0305E82C3301

- 16 bytes
- Universally unique
- Many per table

MORE TERMS

Numeric

Integer

BIG Integer

Small number

MORE TERMS

Clustered Key

Partitioning Key

MORE TERMS

- Index
- Constraint
- Application code
- Referential Integrity

HOW EASY IS THIS?

“Just allocate a surrogate key for every table”

Job done.

And why do we need data modelers/designers?

WHAT YOU SHOULD DO:

- Read Up
- Ensure you use correct terms
- Understand how your tools create and generate keys
- Learn about the Outliers
- Set the standard for correct term use

THANK YOU, YOU WERE GREAT.

KAREN LOPEZ @DATACHICK