Fairfield University
DOLAN
School of Business

# Databases for Analytics

## Assorted SQL Best Practices

How not to bork everything and have to start over

# Tip #1: Choose names carefully

- Consistency is key to avoiding bugs
- Name your entities for the THINGS being tracked, not the containers we put them in
- If using plurals for table names, then do so for every table
- Each attribute should be **singular**
- Consider reserving the *_id naming pattern for surrogate keys only
- **Don't make up terminology! Use names from the domain.**

# Tip #2: Column Order Matters ...

- Follow a repeatable pattern, perhaps something like …
  - PK fields
  - Alternate key fields
  - Regular attributes
  - FK fields
- This makes it trivial to define and check your FK fields
- Also, much easier to read

# Tip #3: Take care with your grammar

- In English we learned about nouns and verbs. The nouns are **things** while verbs are **actions**.
- In data modeling … we have **entities** (nouns), **attributes** (nouns), and **relationships** (verbs)
- When considering a noun in your domain, it can represent either an entity or an attribute value. How can you tell?
  - Entities always have unique identifier attributes
  - Anything else is just an attribute or the value of an attribute

# Tip #4: Define a repeatable build process

- It should **always** be possible to recreate everything from the original source data
- Treat the build process as **more important** than the resulting dataset
- **Document the heck out of the process** so you know exactly what each step does. Or better, so that the next person know what each step does.

# Tip #5: Keep ERD and Code in Sync

- The ERD is useful when
  - creating the tables
  - populating the tables
  - querying the database

  So, that's basically, forever and always!
- The same goes for the data dictionary

**Treat these things like critical code or data**

# Tip #6: Keep simple stuff simple

- When creating or populating tables, always work 'inward' from the strong entities:
  - 1. Strong entities
  - 2. Entities that only refer to strong entities (1)
  - 3. Entities that only refer to entities in 1 and 2
  - …
- If you find that you can't keep track of what's already been decided/built, then how can you expect anybody else to?

# Tip #7: Write queries incrementally

- When writing INSERT queries with SELECTs, always write (and check) the SELECTs first
- When writing SELECTs, always get the JOINs right before worrying about the attributes
- When writing chained JOINs, add and test each JOIN one at a time
- ...

**This makes it much easier to debug!**

# Tip #8: Test early and often

- Before adding anything new to your database, always test/check what exists first.
  - Are the PKs right?
  - Are the datatypes the same on both sides of each FK → PK relationship?
  - Has the data loaded been loaded correctly so far?
- If you find a bug, fix it before moving on. You may need to update the ERD, SQL DDL, and SQL DML.

Fairfield University
DOLAN
School of Business

# Databases for Analytics

## Assorted SQL Best Practices

How not to bork everything and have to start over