

ABHIJIT AGARWAL / DATA SCIENCE RETREAT

# CONCEPTS OF RELATIONAL DATABASES

## DATABASES

# TYPES OF DATABASES

C1	C2	C3	C4
—	—	—	—
—	—	—	—
—	—	—	—
—	—	—	—
—	—	—	—

### Relational data model

Highly-structured table organization with rigidly-defined data formats and record structure.



### Document data model

Collection of complex documents with arbitrary, nested data formats and varying “record” format.

# WHY DATABASES?

- Security - Safety, Backup and Recovery
- Consistency
- Complex Queries - View, Update
- Fast and Efficient - Indexing
- Multiple Processes -Transactions
- Different UIs
- Interactive
- Integrity - Redundancy, Uniqueness, Relations, Data Types

ADVANTAGES

# SECURITY

Authenticity ensured

Passwords

ADVANTAGES

# CONCURRENCY



# ADVANTAGES

# CONSISTENCY

Geography Class

STUDENT NAME	ENROLLED SINCE	EMAIL
MARK	01/07/2014	<u>MARK@DOMAIN.COM</u>
LARS	01/08/2014	<u>LARS@DOMAIN.COM</u>

Name	Id	Email	Dob
Mark	11	<u>Mark@Domain.Com</u>	29/02/1992
Lars	12	<u>Lars@Domain.Com</u>	07/08/1991
Mary	13	<u>Mary@Domain.Com</u>	19/06/1992

Student Table

Maths Class

Student Name	Enrolled Since	Email
Mark	01/08/2014	<u>MARK@DOMAINTN.COM</u>
Mary	15/06/2014	<u>MARY@DOMAINTN.COM</u>

## ADVANTAGES

# CONSISTENCY

Name	Id	Email	Dob
Mark	11	<u>Mark@Domain.Com</u>	29/02/1992
Lars	12	<u>Lars@Domain.Com</u>	07/08/1991
Mary	13	<u>Mary@Domain.Com</u>	19/06/1992



Name	Id	Email	Dob
Mark	11	<u>Mark@Host.Com</u>	29/02/1992
Lars	12	<u>Lars@Domain.Com</u>	07/08/1991
Mary	13	<u>Mary@Domain.Com</u>	19/06/1992

# ADVANTAGES

# CONSISTENCY

Name	Id	Email	Dob
Mark	11	<a href="mailto:Mark@Host.Com">Mark@Host.Com</a>	29/02/1992
Lars	12	<a href="mailto:Lars@Domain.Com">Lars@Domain.Com</a>	07/08/1991
Mary	13	<a href="mailto:Mary@Domain.Com">Mary@Domain.Com</a>	19/06/1992

Student Table

STUDENT NAME	ENROLLED SINCE	EMAIL
MARK	01/07/2014	<a href="mailto:MARK@HOST.COM">MARK@HOST.COM</a>
LARS	01/08/2014	<a href="mailto:LARS@DOMAIN.COM">LARS@DOMAIN.COM</a>

Maths Class

Student Name	Enrolled Since	Email
Mark	01/08/2014	<a href="mailto:MARK@HOST.COM">MARK@HOST.COM</a>
Mary	15/06/2014	<a href="mailto:MARY@DOMAIN.COM">MARY@DOMAIN.COM</a>

ADVANTAGES

# COMPLEX QUERIES

Education Level	Good	Pearl	Portsmouth	Top Measure	Walrus	Good	Pearl	Portsmouth	Top Measure	Walrus
Bachelors Degree	\$65.99	\$107.45	\$132.21	\$33.84	\$139.20	\$402.52	\$213.41	\$414.85	\$269.75	\$371.17
Graduate Degree	\$6.39	\$8.73	\$42.66	\$8.28	\$41.15	\$74.43	\$62.25	\$63.53	\$55.97	\$119.70
High School Degree	\$103.36	\$107.95	\$251.90	\$63.33	\$193.00	\$395.68	\$312.97	\$454.03	\$331.91	\$398.73
Partial College	\$28.48	\$30.00	\$59.04	\$20.27	\$30.07	\$92.07	\$71.97	\$182.36	\$73.84	\$117.90
Partial High School	\$97.76	\$77.42	\$169.12	\$75.39	\$159.35	\$378.87	\$286.88	\$491.04	\$343.83	\$420.37

Dummy Data

Profit per Brand per Education Level in a Supermarket

ADVANTAGES

# SPEED AND EFFICIENCY



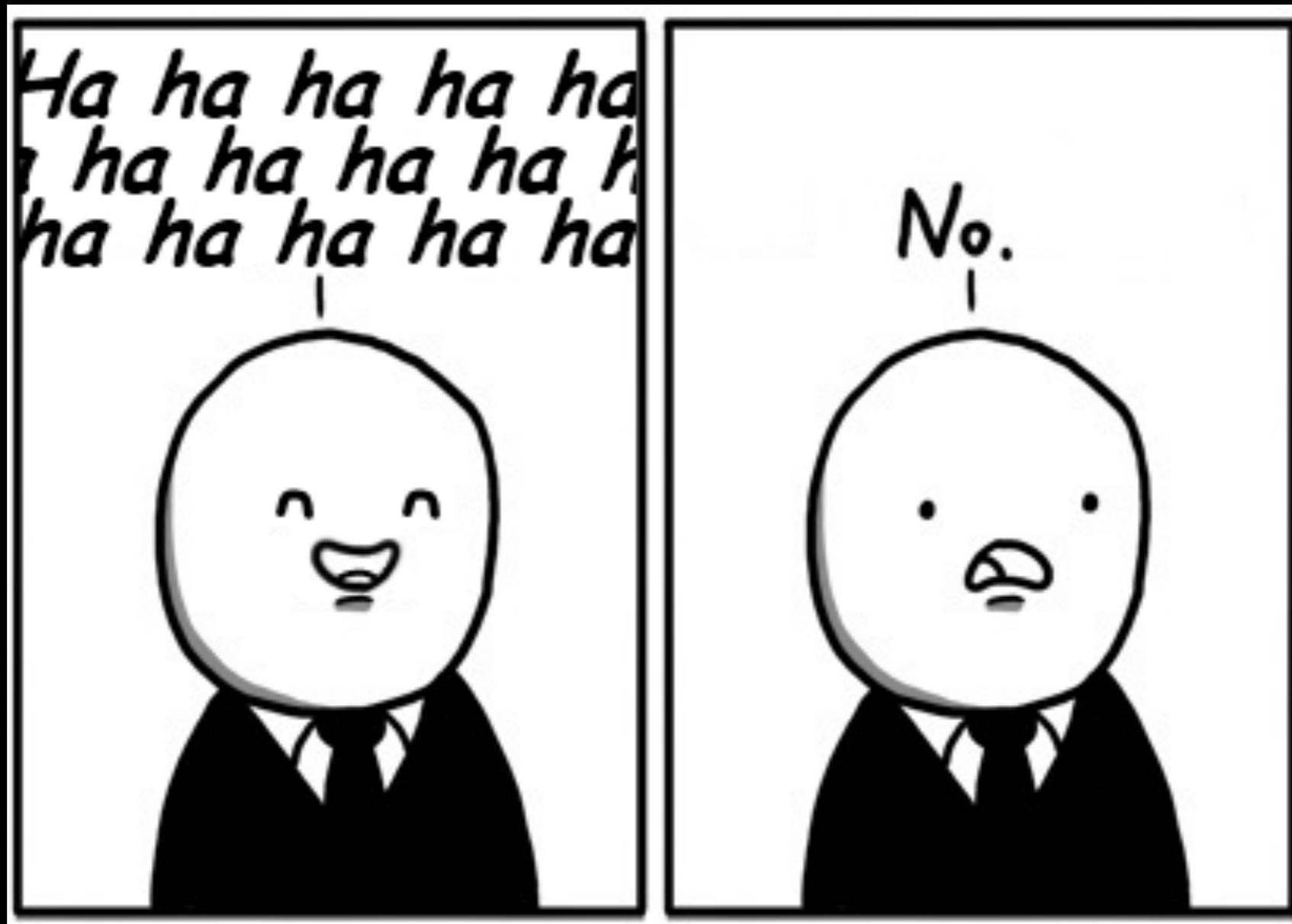
ADVANTAGES

# USER INTERFACES



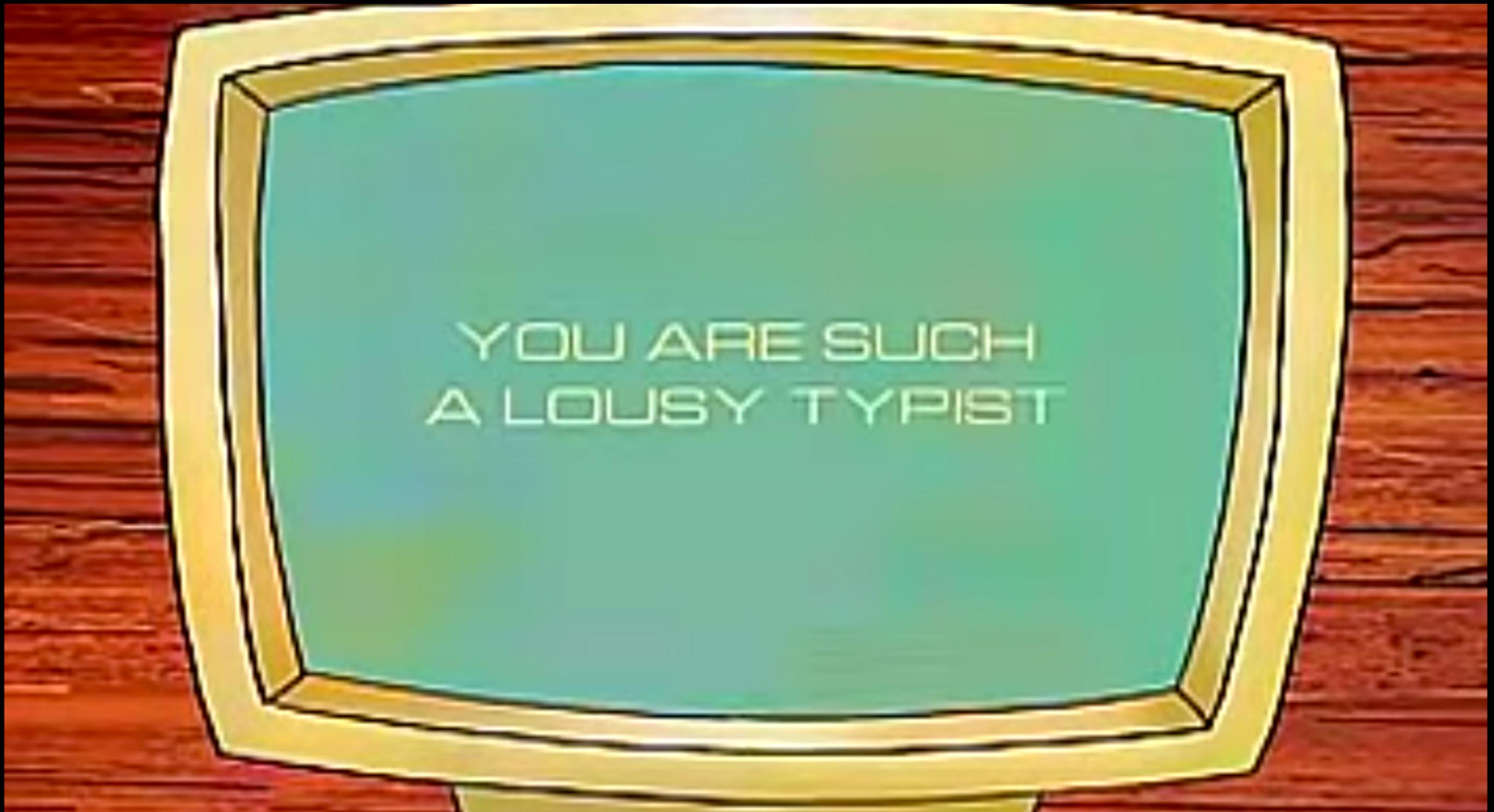
ADVANTAGES

# RESTRICTIVE



ADVANTAGES

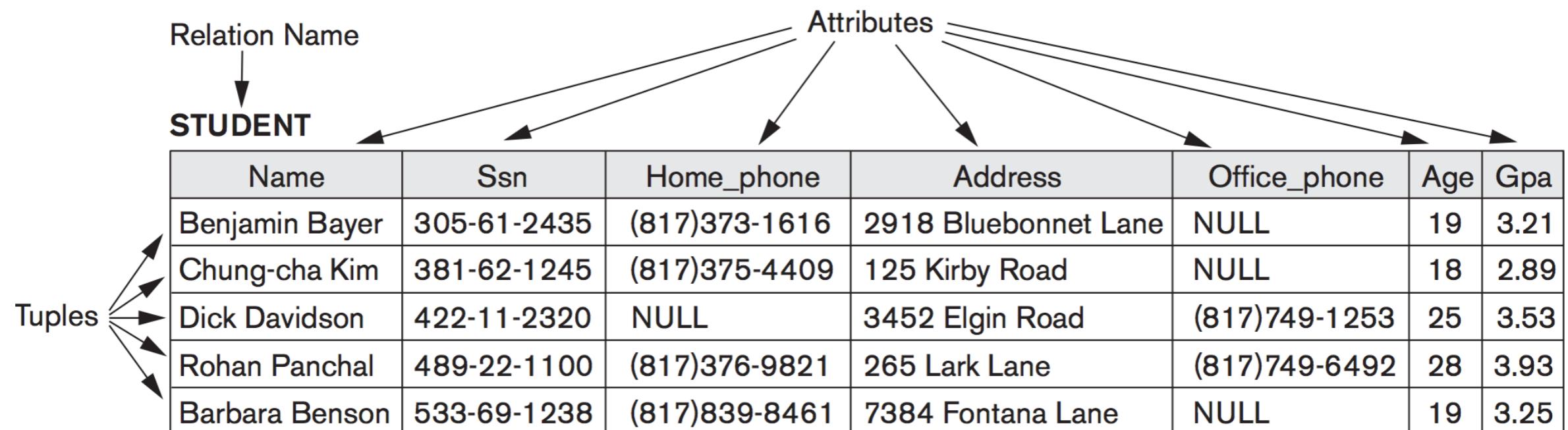
# INTERACTIVE

A stack of books is shown on a dark brown wooden shelf. The books have yellow spines and green covers. The top book's cover features the text "YOU ARE SUCH A LOUSY TYPIST" in white, bold, sans-serif capital letters.

YOU ARE SUCH  
A LOUSY TYPIST

# RELATIONAL DATABASES

# RELATION



The Attributes and Tuples of relation STUDENT

# RELATIONAL DATABASES

## EXERCISE 1:

- Create a relation (table definition) that could be used to store information about an Instagram User
- Just the User - Nothing else (no posts / followers / etc.)
- As simple as possible, please. Not too many columns.



theawkwardyeti ✅

Following



...

1,425 posts

1.7m followers

157 following

**The Awkward Yeti (Nick Seluk)**

How I Broke Up With My Colon, Available for preorder on Amazon, BN.com and Indiebound!

[theawkwardstore.com](http://theawkwardstore.com)

# KEYS

- Candidate Keys
- Primary Key
- Foreign Key



# THE NULL VALUE

- NULL does not mean zero.
- NULL does not mean "" (blank)
- NULL means nothing
- When is null used?
  - When an attribute is not applicable for a record
  - When value is not (yet) known

# DATABASE SCHEMA

- The description of a database is called the database schema, which is specified during database design and is not expected to change frequently.
- A displayed schema is called a schema diagram

# DATABASE EXAMPLE

A Database that stores student and course information

## STUDENT

Name	Student_number	Class	Major
Smith	17	1	CS
Brown	8	2	CS

## COURSE

Course_name	Course_number	Credit_hours	Department
Intro to Computer Science	CS1310	4	CS
Data Structures	CS3320	4	CS
Discrete Mathematics	MATH2410	3	MATH
Database	CS3380	3	CS

## SECTION

Section_identifier	Course_number	Semester	Year	Instructor
85	MATH2410	Fall	07	King
92	CS1310	Fall	07	Anderson
102	CS3320	Spring	08	Knuth
112	MATH2410	Fall	08	Chang
119	CS1310	Fall	08	Anderson
135	CS3380	Fall	08	Stone

## GRADE\_REPORT

Student_number	Section_identifier	Grade
17	112	B
17	119	C
8	85	A
8	92	A
8	102	B
8	135	A

# DATABASE SCHEMA

# SCHEMA OF EXAMPLE

## STUDENT

Name	Student_number	Class	Major
------	----------------	-------	-------

## COURSE

Course_name	Course_number	Credit_hours	Department
-------------	---------------	--------------	------------

## SECTION

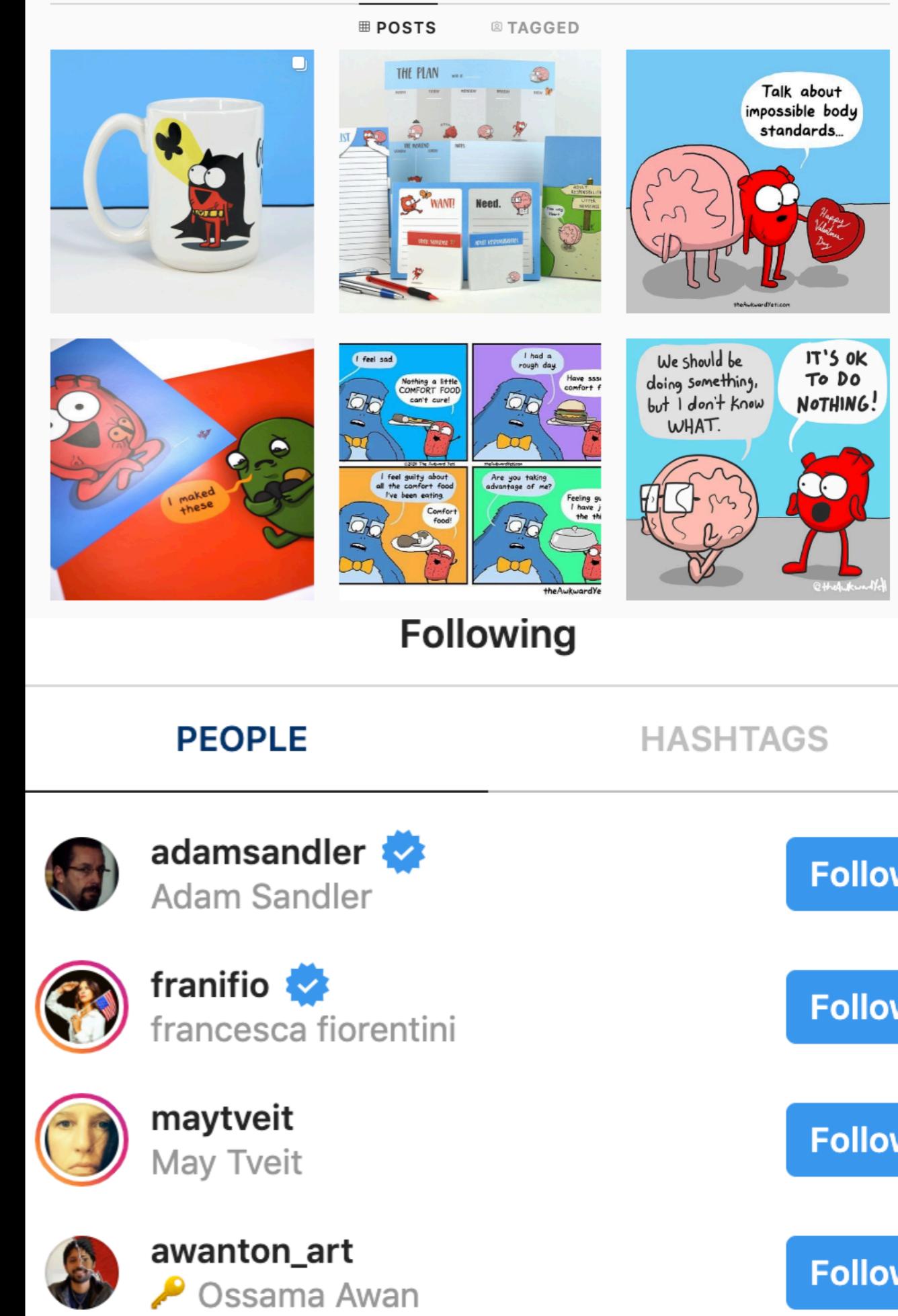
Section_identifier	Course_number	Semester	Year	Instructor
--------------------	---------------	----------	------	------------

## GRADE\_REPORT

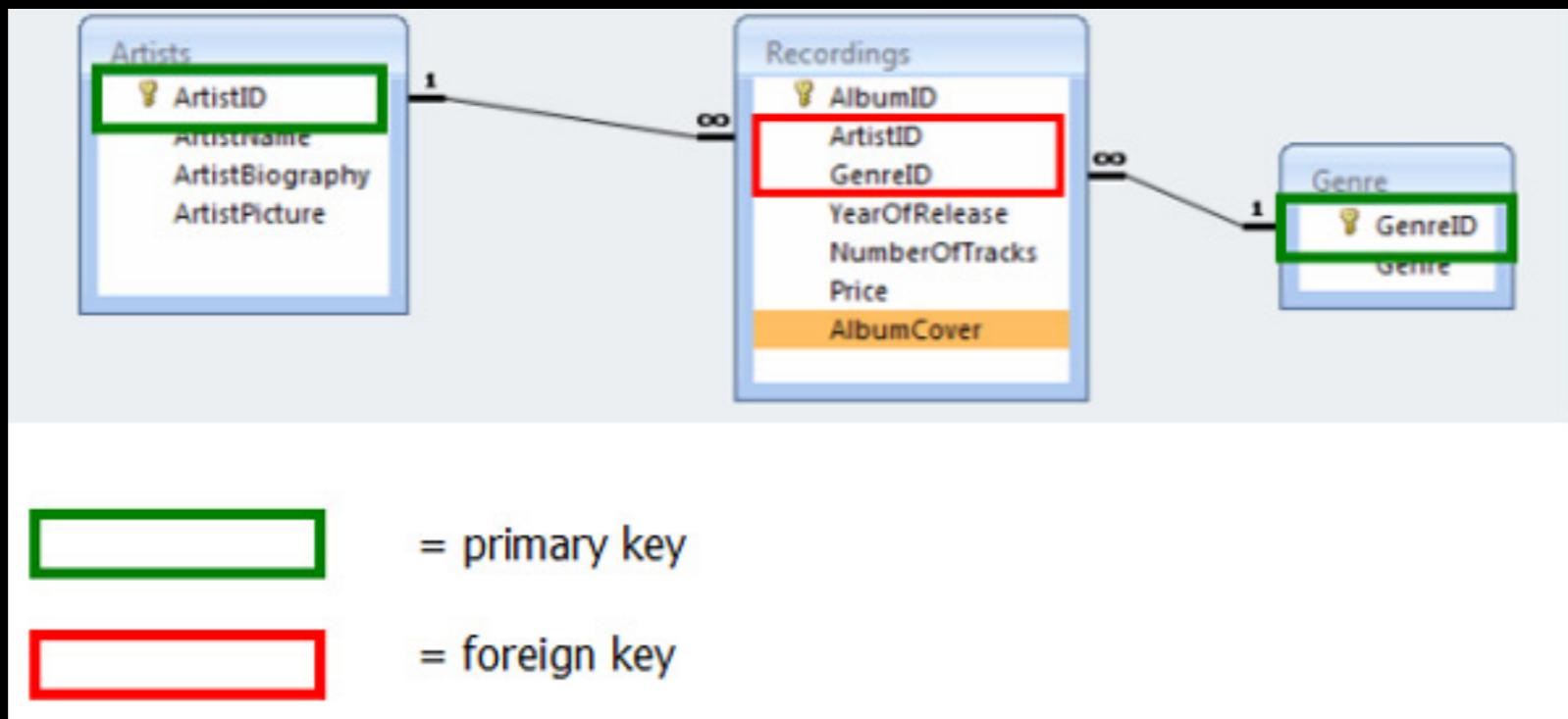
Student_number	Section_identifier	Grade
----------------	--------------------	-------

# RELATIONAL DATABASES EXERCISE 2

- Add a table for Posts
- Add a table for followers and following
- Again, as simple as possible.



# RELATIONSHIPS / FOREIGN KEYS



- One to One Relationship
- One to Many Relationship
- Many to Many Relationship

# CRUD OPERATIONS

- Create - CREATE TABLE, INSERT
- Read - SELECT
- Update - UPDATE
- Delete - DROP TABLE, DELETE

# JOINS

## Orders Table

OrderID	CustomerID	OrderDate
10308	2	1996-09-18
10309	37	1996-09-19
10310	77	1996-09-20

# JOINS

## Orders Table

OrderID	CustomerID	OrderDate
10308	2	1996-09-18
10309	37	1996-09-19
10310	77	1996-09-20

## Customers Table

CustomerID	CustomerName	ContactName	Country
1	Alfreds Futterkiste	Maria Anders	Germany
2	Ana Trujillo Emparedados y helados	Ana Trujillo	Mexico
3	Antonio Moreno Taquería	Antonio Moreno	Mexico

# JOINS

OrderID	CustomerID	OrderDate	
10308	2	1996-09-18	
10309	37	1996-09-19	
10310	77	1996-09-20	
CustomerID	CustomerName	ContactName	Country
1	Alfreds Futterkiste	Maria Anders	Germany
2	Ana Trujillo Emparedados y helados	Ana Trujillo	Mexico
3	Antonio Moreno Taquería	Antonio Moreno	Mexico

OrderID	CustomerName	OrderDate
10308	Ana Trujillo Emparedados y helados	9/18/1996
10365	Antonio Moreno Taquería	11/27/1996
10383	Around the Horn	12/16/1996
10355	Around the Horn	11/15/1996
10278	Berglunds snabbköp	8/12/1996

# TYPES OF JOINS

- Inner
- Left or Left Outer
- Right or Right Outer
- Outer or Full Outer
- Semi: Left or Right
- Anti: Left or Right
- Cross Join



# CONCEPTS OF RELATIONAL DATABASE NORMALIZATION

Book_Id	Book Title	Author	Category	Format	Price	Publisher	Publisher Country
1	Lethal White	Robert Galbraith	FICTION, CRIME	Paperback	10,99	Penguin	Uk
2	In Cold Blood	Truman Capote	NON-FICTION, CLASSICS, CRIME	Kindle	7,99	Puffin	Uk
2	In Cold Blood	Truman Capote	NON-FICTION, CLASSICS, CRIME	Paperback	9,99	Puffin	Uk

# NORMALIZATION: 1 NF

Dealing with multi-value columns

Book_Id	Book Title	Author	Format	Price	Publisher	Publisher Country
1	Lethal White	Robert Galbraith	Paperback	10,99	Penguin	Uk
2	In Cold Blood	Truman Capote	Kindle	7,99	Puffin	Uk
2	In Cold Blood	Truman Capote	Paperback	9,99	Puffin	Uk

Category_Id	Category Name
1	Fiction
2	Non-Fiction
3	Crime

Book_Id	Category_Id
1	1
1	3
2	2
2	3

# 2 N F

Don't repeat information for one column

Book_Id	Book Title	Author	Publisher	Publisher Country
1	Lethal White	Robert Galbraith	Penguin	Uk
2	In Cold Blood	Truman Capote	Puffin	Uk

Format_Id	Format Name
1	Paperback
2	Harcover
3	Kindle

Book_Id	Format_Id	Price
1	1	10,99
2	1	7,99
2	3	9,99

# 3 N F

No transitive dependencies / correlation

Book_Id	Book Title	Author	Publisher_Id
1	Lethal White	Robert Galbraith	1
2	In Cold Blood	Truman Capote	2

Publisher_Id	Publisher	Publisher_Country
1	Penguin	Uk
2	Puffin	Uk

# PROS AND CONS

## Pros

Data Updates are atomic

Data Updates are consistent

Takes less storage space, due to less repetitions

## Cons

Querying speed is reduced, too many joins

A lot more complicated to design and operate

# RELATIONAL DATABASES

## EXERCISE 3

Create normalized tables for a articles on a news website

- Article Title ("10 things that prove we are living in The Matrix")
- Article Body ("...")
- Author Name ("Keanu Reeves")
- Author Bio ("Keanu is a veteran Actor")
- Tags ("The Matrix", "Keanu Reeves", "Pop Culture")
- Time of Post (2020-02-29 04:20:00)
- Comments ("Max, max@emailprovider: Mr. Reeves is a Legend")