॥ त्वं ज्ञानमयो विज्ञानमयोऽसि ॥
IIT Jodhpur

Excellence above all
SME

Technology • Innovation • Leadership

The (Small/Relational) Database Approach

Dr. Deepak Saxena, SME IIT Jodhpur

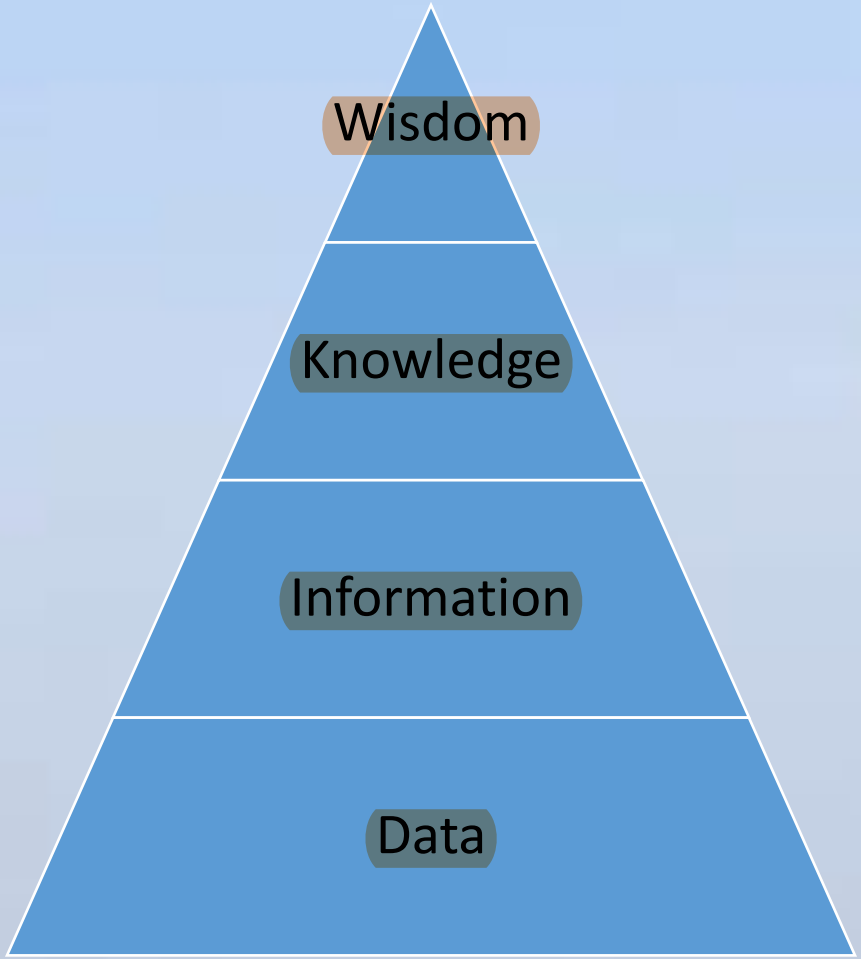

DATA
IS THE NEW OIL

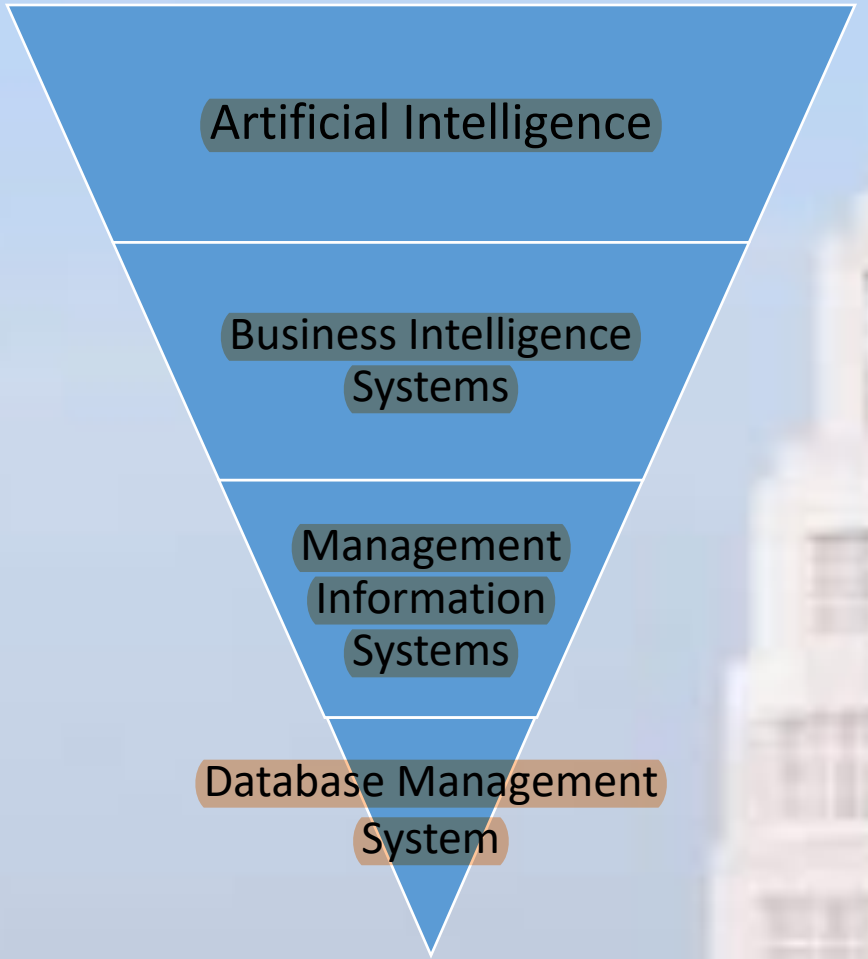

"Guide us, Oh Database Manager!"

Data, Information, and Metadata

- **Data**: Stored representations of objects and events that have meaning and importance in the users' environment.
  - *Objects*: Customer, Items
  - *Events*: Booking, Payment, Delivery
- **Information**: Data that have been processed (e.g. averages, trends etc.) in such as way as to increase the knowledge of the person who uses the data.
- **Metadata**: Data that describes the properties of the end-user data and the context of that data.
- **Database**: an organized collection of logically-related data

Information Systems for DIKW Model





Traditional File Processing Systems

- Program-file dependence
- Duplication of data
- Limited data sharing
- Lengthy development times
- Excessive program maintenance

Customer	Meal1	Date1	Cost1	Meal2	Date2	Cost2
J. Smith	Steak	2/1/2013	$ 20.00	Lobster	2/3/2013	$ 25.00
Jan						
R. Doyle	Veal	3/1/2013	$ 30.00	Shrimp	5/10/2013	$ 20.00
E. Pengler	Steak	2/5/2013	$ 20.00	Steak	7/8/2013	$ 20.00

Relational Data Model

- The relational data model represents data in the form of tables.
- Consists of the following three components:
  1. ***Data structure*** Data are organized in the form of tables, with rows and columns.
  2. ***Data manipulation*** Powerful operations (typically implemented using SQL) are used to manipulate data stored in the relations.
  3. ***Data integrity*** The model includes mechanisms to specify business rules that maintain the integrity of data when they are manipulated.

Relation

- A named, two-dimensional table of data.
- Each relation (or table) consists of a set of named columns and an arbitrary number of unnamed rows.
- An attribute is a named column of a relation.
- Each row of a relation corresponds to a record that contains data (attribute) values for a single entity.
- Example: EMPLOYEE1(EmpID, Name, DeptName, Salary)

Properties of Relations

1. Each relation (or table) in a database has a unique name.
2. An entry at the intersection of each row and column is atomic (or single valued).There can be only one value associated with each attribute on a specific row of a table; no multivalued attributes are allowed in a relation.
3. Each row is unique; no two rows in a relation can be identical.
4. Each attribute (or column) within a table has a unique name.
5. The sequence of columns (left to right) is insignificant. The order of the columns in a relation can be changed without changing the meaning or use of the relation.
6. The sequence of rows (top to bottom) is insignificant. As with columns, the order of the rows of a relation may be changed or stored in any sequence.

No multivalued attributes in a relation

(a) Table with repeating groups

EmpID	Name	DeptName	Salary	CourseTitle	DateCompleted
100	Margaret Simpson	Marketing	48,000	SPSS	6/19/2015
				Surveys	10/7/2015
140	Alan Beeton	Accounting	52,000	Tax Acc	12/8/2015
110	Chris Lucero	Info Systems	43,000	Visual Basic	1/12/2015
				C++	4/22/2015
190	Lorenzo Davis	Finance	55,000		
150	Susan Martin	Marketing	42,000	SPSS	6/16/2015
				Java	8/12/2015

(b) EMPLOYEE2 relation

EMPLOYEE2

EmpID	Name	DeptName	Salary	CourseTitle	DateCompleted
100	Margaret Simpson	Marketing	48,000	SPSS	6/19/2015
100	Margaret Simpson	Marketing	48,000	Surveys	10/7/2015
140	Alan Beeton	Accounting	52,000	Tax Acc	12/8/2015
110	Chris Lucero	Info Systems	43,000	Visual Basic	1/12/2015
110	Chris Lucero	Info Systems	43,000	C++	4/22/2015
190	Lorenzo Davis	Finance	55,000		
150	Susan Martin	Marketing	42,000	SPSS	6/19/2015
150	Susan Martin	Marketing	42,000	Java	8/12/2015

Anomalies

- An error or inconsistency that may result when a user attempts to update a table that contains redundant data.
- Indication that your table design is not proper.
- Three types:
  - Insertion anomalies
  - Deletion anomaly
  - Modification anomaly

EMPLOYEE2

EmpID	Name	DeptName	Salary	CourseTitle	DateCompleted
100	Margaret Simpson	Marketing	48,000	SPSS	6/19/2015
100	Margaret Simpson	Marketing	48,000	Surveys	10/7/2015
140	Alan Beeton	Accounting	52,000	Tax Acc	12/8/2015
110	Chris Lucero	Info Systems	43,000	Visual Basic	1/12/2015
110	Chris Lucero	Info Systems	43,000	C++	4/22/2015
190	Lorenzo Davis	Finance	55,000		
150	Susan Martin	Marketing	42,000	SPSS	6/19/2015
150	Susan Martin	Marketing	42,000	Java	8/12/2015

Solution: Normalization

EmpID	CourseTitle	DateCompleted
100	SPSS	6/19/2015
100	Surveys	10/7/2015
140	Tax Acc	12/8/2015
110	Visual Basic	1/12/2015
110	C++	4/22/2015
150	SPSS	6/19/2015
150	Java	8/12/2015

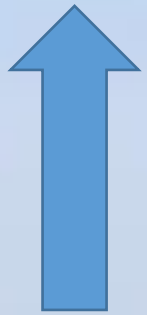
Relational Keys

- **Primary key:** An attribute or a combination of attributes that uniquely identifies each row in a relation.
- Example: EMPLOYEE1(EmpID, Name, DeptName, Salary)

- **Composite key:** A primary key that consists of more than one attribute.
- Example: DEPENDENT1 (EmpID, DependentName, Relationship)

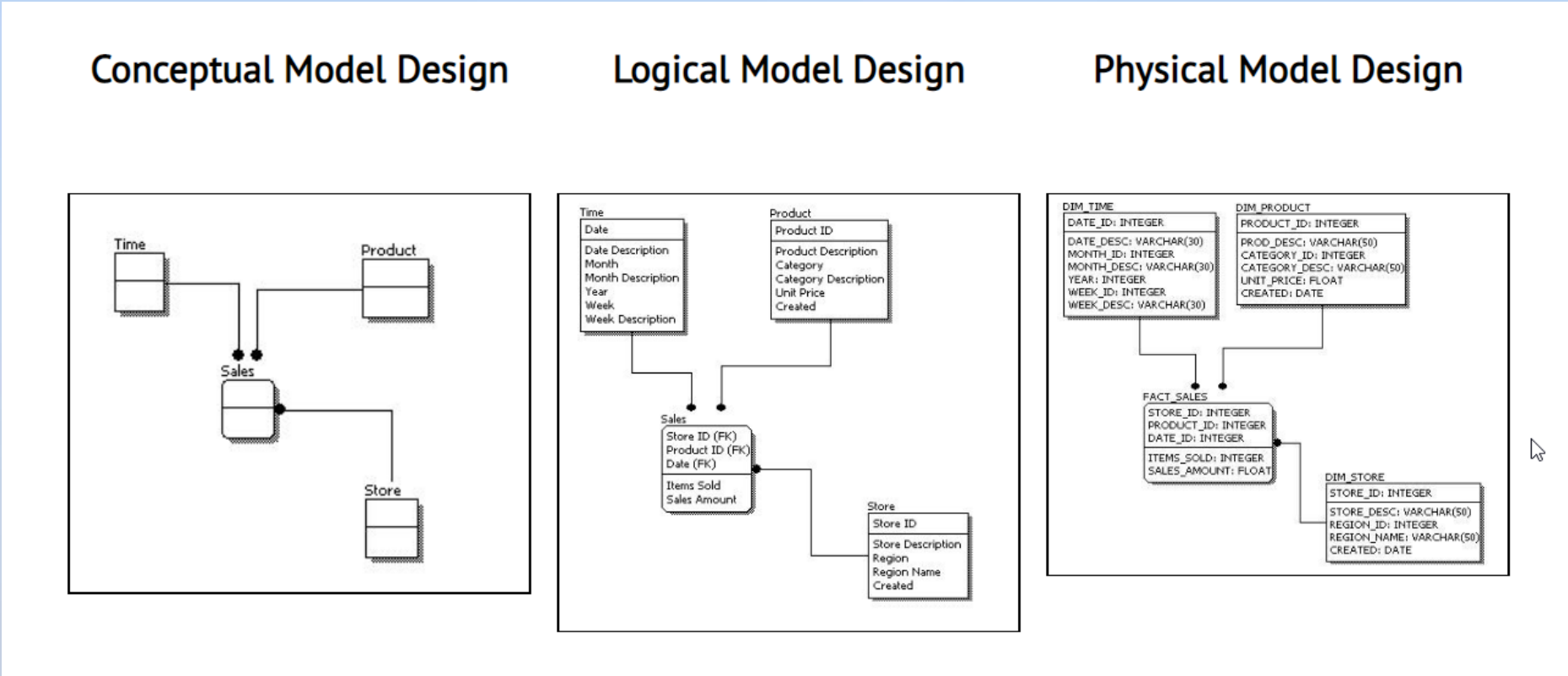
Relational Keys

- **Foreign key:** An attribute in a relation that serves as the primary key of another relation in the same database.
- DEPARTMENT(DeptID, DeptName, Location, Fax)
- EMPLOYEE1(EmpID, Name, DeptID, Salary)

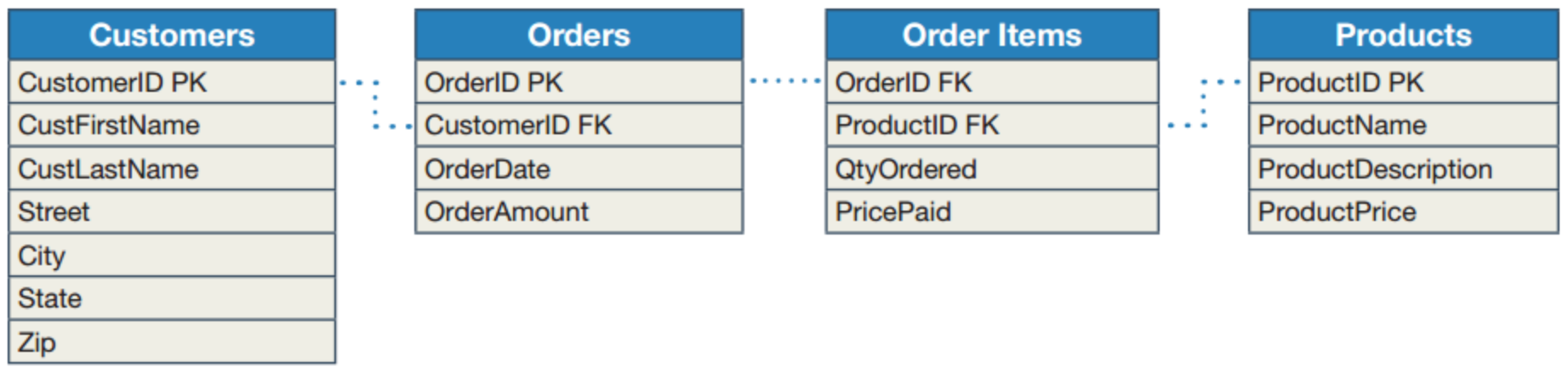


Foreign Key

- The database approach: Data models



- The database approach: Relational database



Integrity Constraints

- Domain Constraints
  - All of the values that appear in a column of a relation must be from the same domain.
  - A domain is the set of values that may be assigned to an attribute.
- Entity Integrity
  - No primary key attribute (or component of a primary key attribute) may be null.
- Referential Integrity
  - Either each foreign key value must match a primary key value in another relation or the foreign key value must be null.

Domain Constraints

- A domain definition usually consists of the following components: domain name, meaning, data type, size (or length), and allowable values or allowable range (if applicable).

TABLE 4-1 Domain Definitions for INVOICE Attributes

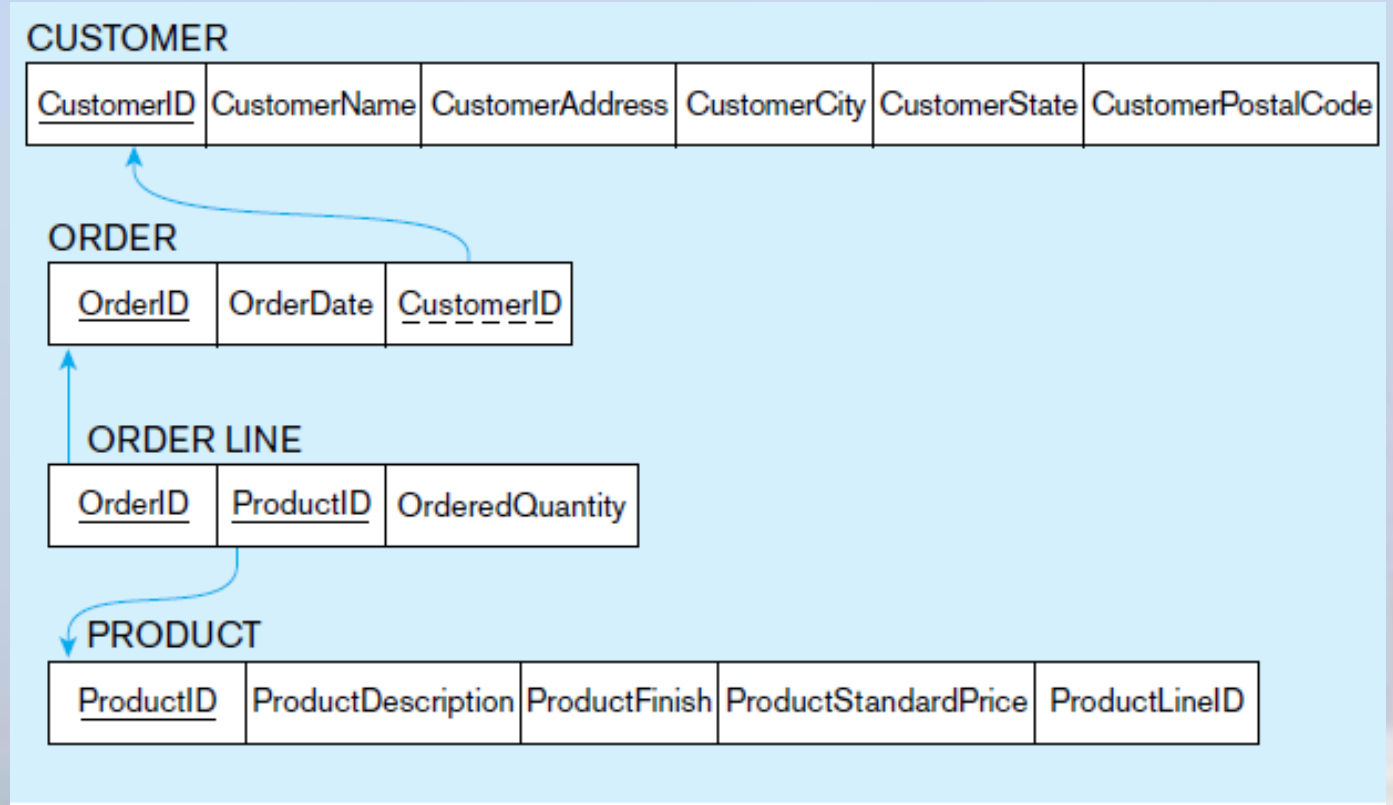
Attribute	Domain Name	Description	Domain
CustomerID	Customer IDs	Set of all possible customer IDs	character: size 5
CustomerName	Customer Names	Set of all possible customer names	character: size 25
CustomerAddress	Customer Addresses	Set of all possible customer addresses	character: size 30
CustomerCity	Cities	Set of all possible cities	character: size 20
CustomerState	States	Set of all possible states	character: size 2
CustomerPostalCode	Postal Codes	Set of all possible postal zip codes	character: size 10
OrderID	Order IDs	Set of all possible order IDs	character: size 5
OrderDate	Order Dates	Set of all possible order dates	date: format mm/dd/yy
ProductID	Product IDs	Set of all possible product IDs	character: size 5
ProductDescription	Product Descriptions	Set of all possible product descriptions	character: size 25
ProductFinish	Product Finishes	Set of all possible product finishes	character: size 15
ProductStandardPrice	Unit Prices	Set of all possible unit prices	monetary: 6 digits
ProductLineID	Product Line IDs	Set of all possible product line IDs	integer: 3 digits
OrderedQuantity	Quantities	Set of all possible ordered quantities	integer: 3 digits

Entity Integrity

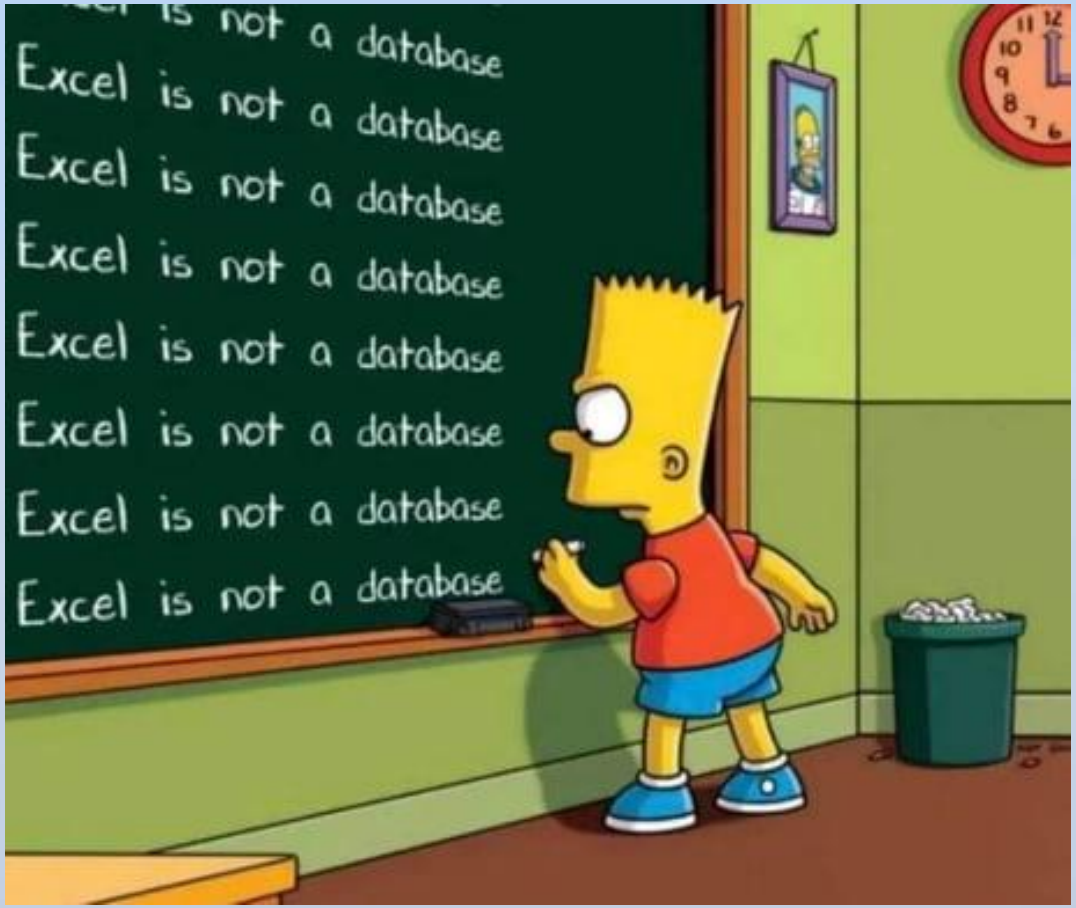
- The entity integrity rule is designed to ensure that every relation has a primary key and that the data values for that primary key are all valid.
- In particular, it guarantees that every primary key attribute is non-null.
- Null Values
  - A null is a value that may be assigned to an attribute when no other value applies or when the applicable value is unknown.
  - In reality, a null is not a value, but rather it indicates the absence of a value.
  - The inclusion of nulls in the relational model is somewhat controversial, because it sometimes leads to anomalous results.
  - However, Codd, the inventor of the relational model, advocates the use of nulls for missing values.

Referential Integrity

- Referential integrity constraint is a rule that maintains consistency among the rows of two relations.
- If there is a foreign key in one relation, either each foreign key value must match a primary key value in another relation or the foreign key value must be null.



Hence…


is not a database
Excel is not a database
Excel is not a database
Excel is not a database
Excel is not a database
Excel is not a database
Excel is not a database
Excel is not a database
Excel is not a database