Database Storage Models

How is data stored in a database?

- Different types of models
  - Conceptual: what is stored, not how
  - Logical: how data is organized
  - Physical: how data is stored physically

- Examples from

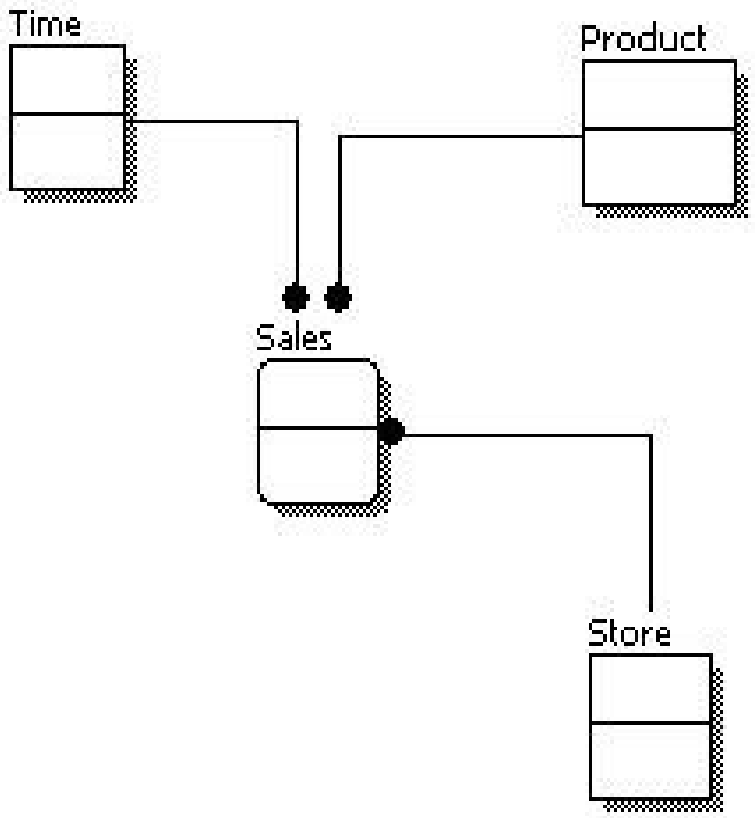
https://www.1keydata.com/datawarehousing

Conceptual Model

- A conceptual data model identifies the highest-level relationships between the different entities.

  Features of conceptual data model include:
  - Important entities and their relationships.
  - No attribute is specified.
  - No primary key is specified.

Conceptual Model Example



In this example, Sales depends on Time, Product, and Store. However, we don't know much else. We just know they are connected.

Logical Model

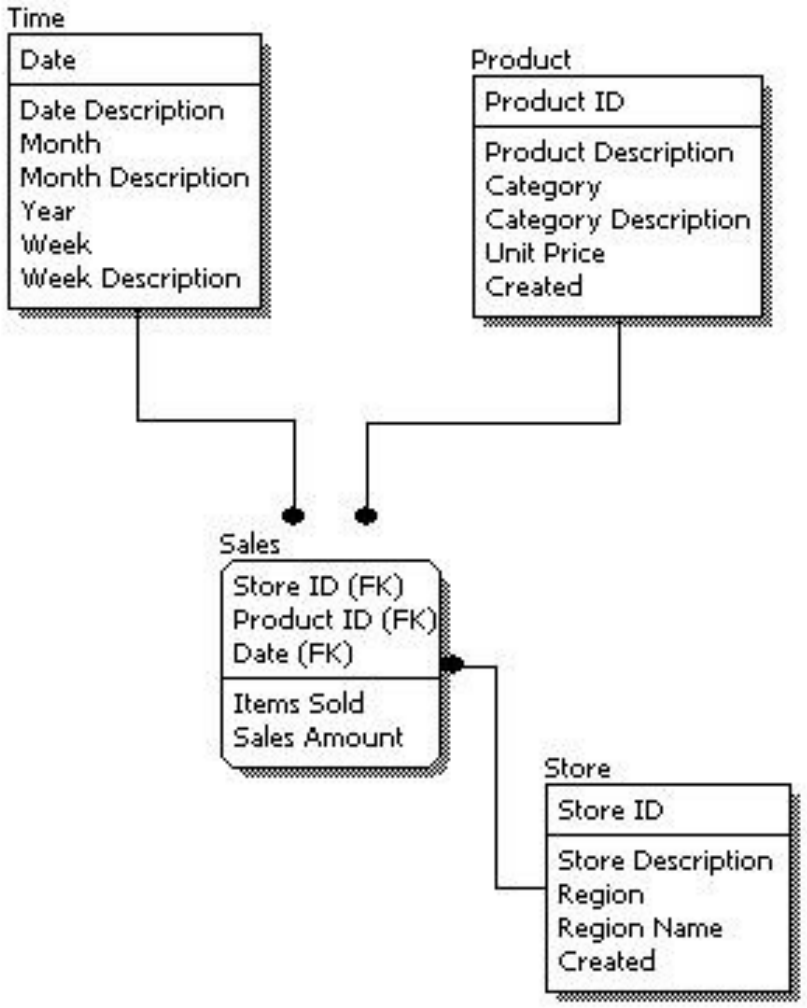
- A logical data model describes the data in as much detail as possible, without regard to how they will be physical implemented in the database. Features of a logical data model include:

  - All entities and relationships among them.
  - All attributes for each entity are specified.
  - The primary key for each entity is specified.
  - Foreign keys (keys identifying the relationship between different entities) are specified.
  - Normalization occurs at this level.

Logical Model Steps

1. Specify primary keys for all entities.
2. Find the relationships between different entities.
3. Find all attributes for each entity.
4. Resolve many-to-many relationships.
5. Normalization.

Logical Model Example



In this model, we know the attributes, primary keys, and foreign keys.

Keys are above the horizontal dividing lines, and a foreign key (FK) is a copy, or link, of a primary key from another entity.

Physical Model

- A physical database model shows all table structures, including column name, column data type, column constraints, primary key, foreign key, and relationships between tables. Features of a physical data model include:

  - Specification all tables and columns.
  - Foreign keys are used to identify relationships between tables.
  - Denormalization may occur based on user requirements.
  - Physical considerations may cause the physical data model to be different from the logical data model.
  - Physical data model will be different for different **RDBMS**. For example, data type for a column may be different between MySQL and SQL Server.

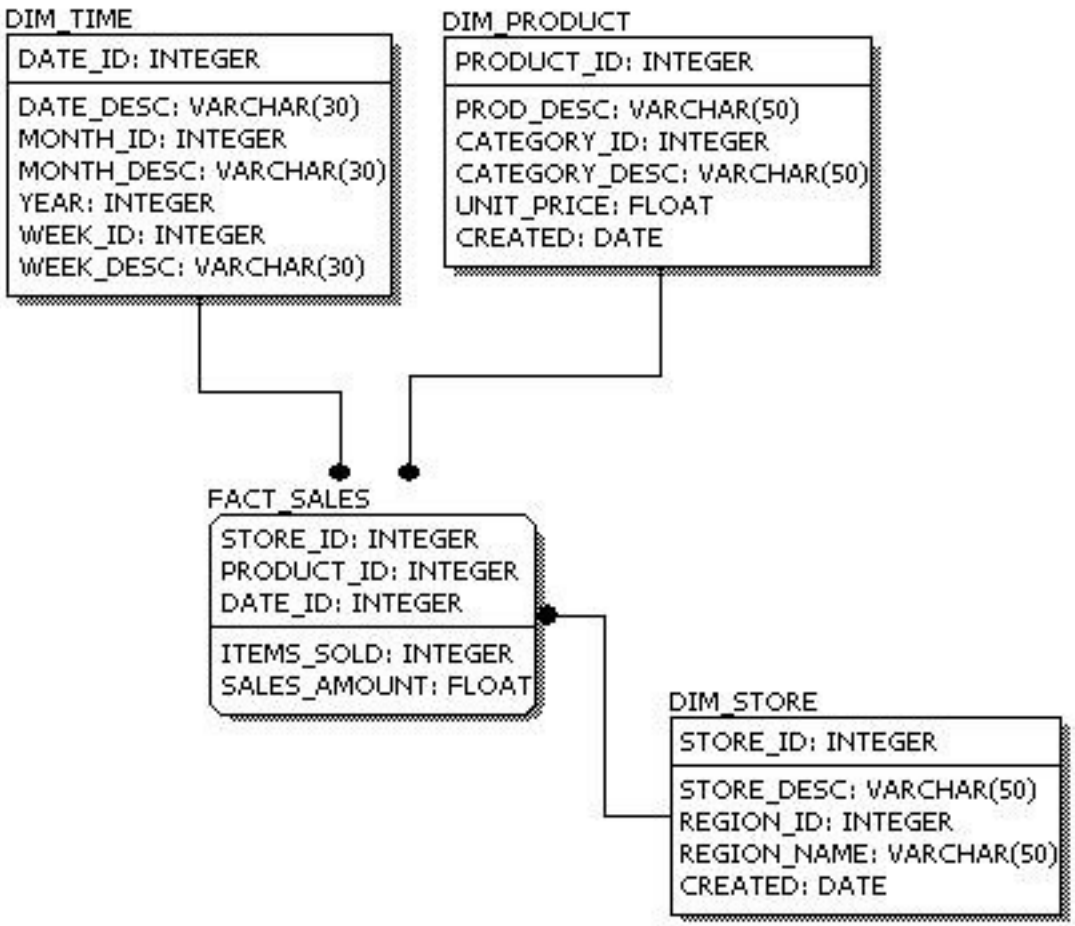
We'll discuss relational databases and other types in a minute

Physical Model

The steps for physical data model design are as follows:

1. Convert entities into tables.
2. Convert relationships into foreign keys.
3. Convert attributes into columns.
4. Modify the physical data model based on physical constraints / requirements.

Physical Model Example

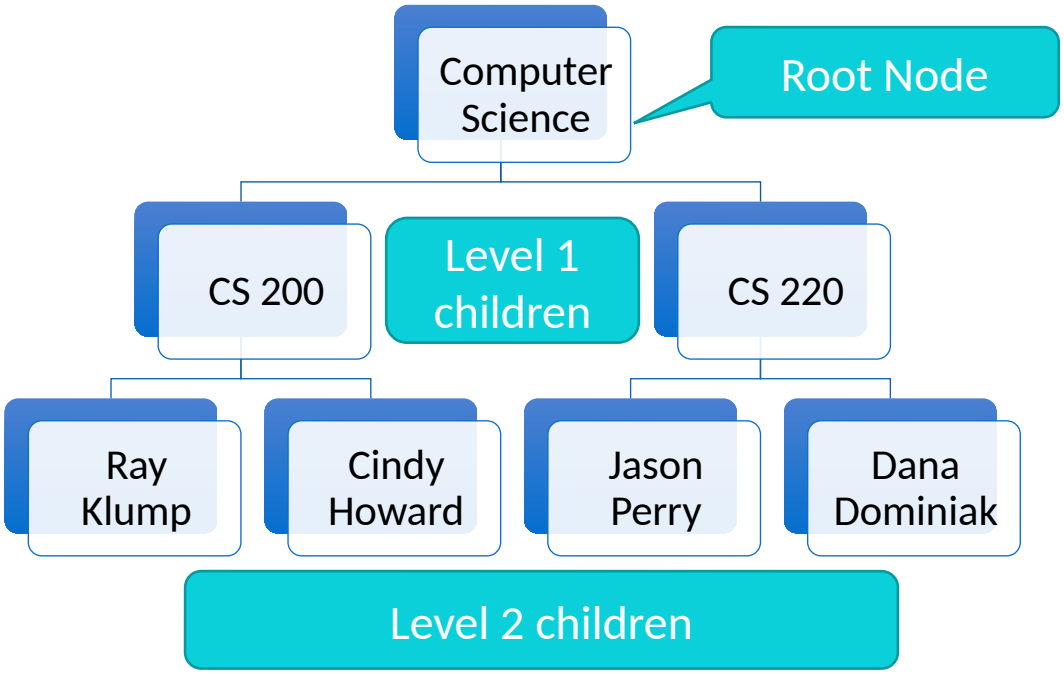


The physical model has the most detail, and describes all names and attributes with software variable names and types.

Storage Models

- First-generation
  - Hierarchical and Network (LDAP)
- Second generation
  - Relational
- Third generation
  - Object-Relational
  - Object-Oriented
- NoSQL (MongoDB)

Hierarchical Database Model Example



LDAP and Hierarchical Databases

- LDAP is the Lightweight Directory Access Protocol
  - Internet ports 389, 636
  - Commands built into the protocol that crosses the Internet
- LDAP supports hierarchical databases (“trees”)
- Includes Open LDAP, *Microsoft Active Directory, Red Hat Directory Server, IBM Tivoli
- Highly efficient for high-reads, low writes

Advantages and Disadvantages of Hierarchical Model

- Advantages
  - Conceptually simple
  - Can be very efficient
  - Ensure data integrity
- Disadvantages
  - Need physical knowledge of how data are stored in order to access it
  - Rigid design so adding a new field requires that the entire database be redesigned
  - Only one parent and no linkages between children
  - No standardized data access language

Network Database Model

- Revolutionary concepts
  - a network schema that the database administrator uses
  - one or more subschemas that provide simpler access to certain sets of nodes
  - Standard languages
- Structure is like the hierarchical model, except that a child can have multiple parents

Advantages and Disadvantages of Network Model

- Advantages. Those of the hierarchical model plus
  - the ability to hold many-to-many relationships
  - standard DML and DDL
  - Navigation is not limited to particular tree traversals
- Disadvantages
  - Navigation - still need to know how the data is organized
  - Structural dependence

The Relational Model

- Introduced in 1970
- Created by E.F. Codd
- Now the standard model for commercial DBMS products
- The relational structure is the most commonly used today.
- It uses two-dimensional rows and columns to store data.
- The tables of records can be connected by common key values.

Relational Model Structure

- Based on mathematical principles
  - Data structure
  - Operations
  - Rules
- Data structure and operations are based on set theory
  - A ***set*** is a collection of values, or elements, with no inherent order
- Relational data structure is based on three mathematical concepts
  - **Domain** is the set of allowable values
  - **Tuple** is a sequence of values
  - A **relation** is a named set of tuples
- **Attribute** is each position in a tuple
  - Attributes are named

Domains

Attribute	Domain name	Meaning	Domain Definition
id	IdNumber	The set of all possible ID numbers at Lewis University	10 digit character string
email	EmailAddress	The set of all possible valid email addresses. Contains a @ followed by a domain.	20 character string
streetAddress	StreetAddress	Set of all valid postal addresses	30 character string
zipCode	ZipCode	Set of all valid zip codes in the United States	10 character string

Relational Model Terminology

Relational Algebra	Database	File System
Domain	Data Type	Data Type
Tuple	Row	Record
Relation	Table	File
Attribute	Column	Field

Relational Operations

- Five basic operations in relational algebra
  - Selection
  - Projection
  - Cartesian product
  - Union
  - Difference
- These perform most of the data retrieval operations needed.
- Also have these operations which can be expressed in terms of 5 basic operations.
  - Join
  - Intersection
  - Division
- All of this means we can do simple or complex selections of various relationships.

Relational Rules

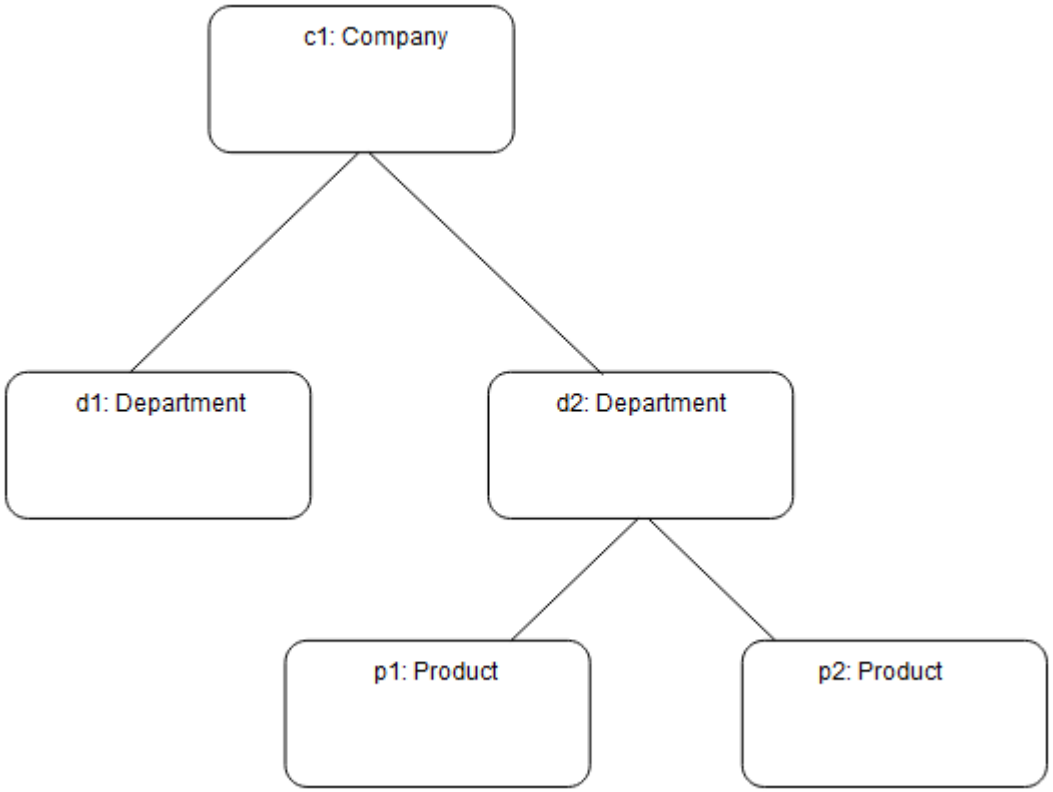
- Integrity rules (more later)
- Structural rules
  - Unique primary key
  - Unique column names
  - No duplicate rows
- Business rules are relational rules specific to a particular database and application.
  - Unique column values
  - No missing values
  - Delete cascade

The Object Model

- Researched in 70's, went 'live' in 1985
- Stores all data as objects – classes, methods, inheritance
- Complex data relationships are fast to query
- Code is more simple
- Simple tasks are inefficient
- Language-dependencies, not widely adopted

Object Model

- MongoDB offers “Realm” as an Object Oriented Database (OOD)
- Uses Java, Javascript, other object-oriented languages for interface



The NoSQL Model

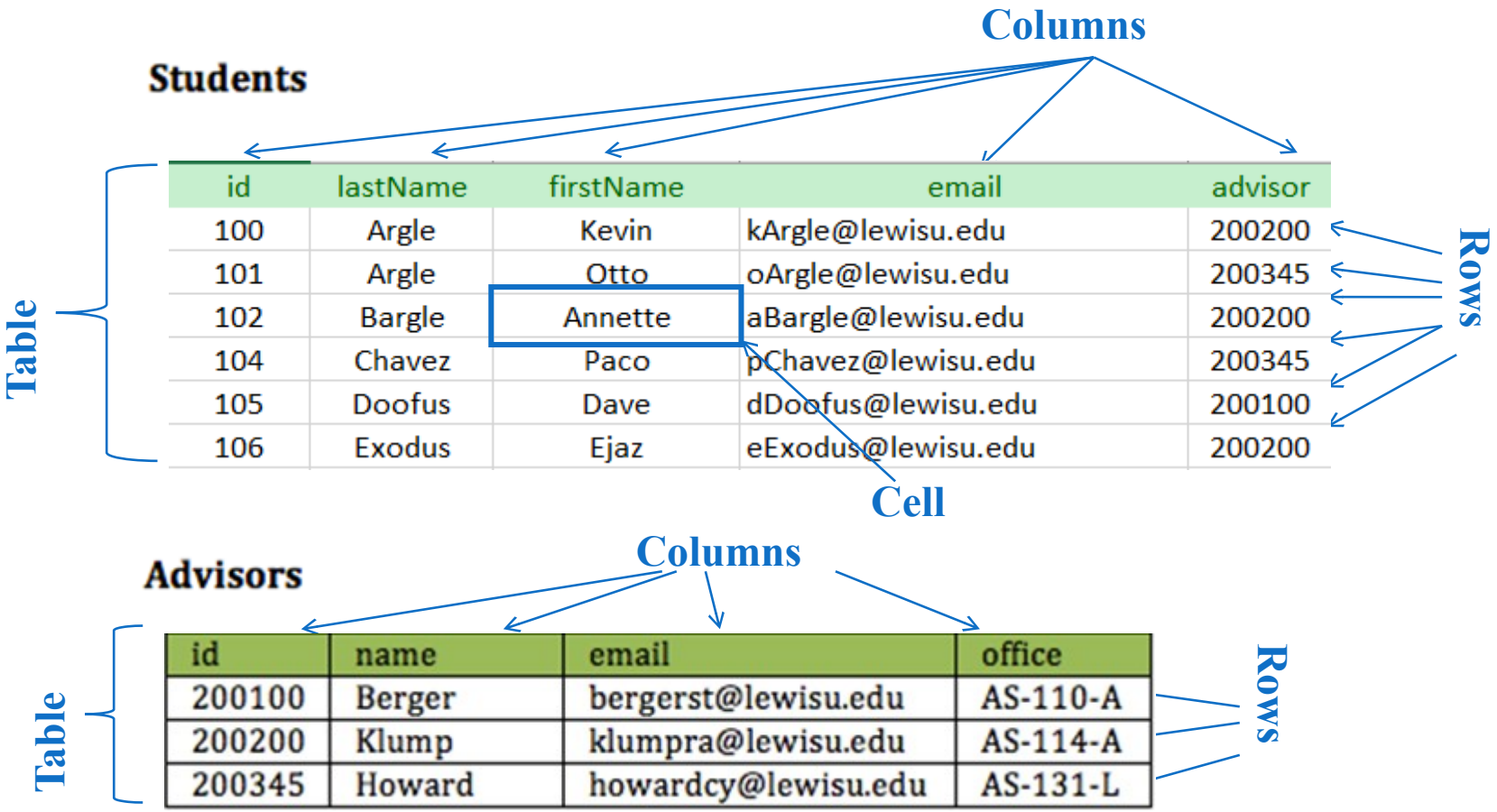
- Means “Not Only SQL”
- 1998- Carlo Strozzi use the term NoSQL for his lightweight, open-source relational database
- Relies on custom scripts and processing power to make sense of the data.
- Loose or no schema, no ACID, no normalizations
- Generally unstructured, very large data applications
- Varieties of storage: Key-value, graph-based, document based

Tables, Columns and Rows

Properties of Tables

- Table name is distinct from all other table names in database.
- Each cell of relation contains exactly one atomic (single) value.
- Each column has a distinct name.
- Each row is distinct; there are no duplicate rows.
- Order of attributes has no significance.
- Order of rows has no significance

Tables, Columns and Rows



id	lastName	firstName	email	advisor
100	Argle	Kevin	kArgle@lewisu.edu	200200
101	Argle	Otto	oArgle@lewisu.edu	200345
102	Bargle	Annette	aBargle@lewisu.edu	200200
104	Chavez	Paco	pChavez@lewisu.edu	200345
105	Doofus	Dave	dDoofus@lewisu.edu	200100
106	Exodus	Ejaz	eExodus@lewisu.edu	200200

id	name	email	office
200100	Berger	bergerst@lewisu.edu	AS-110-A
200200	Klump	klumpra@lewisu.edu	AS-114-A
200345	Howard	howardcy@lewisu.edu	AS-131-L

Data Type Implementations

Category	MySQL	Oracle	SQL Server
Integer	TINYINT SMALLINT MEDIUMINT BIGINT	INT NUMBER	TINYINT SMALLINT INT BIGINT
Decimal	FLOAT DOUBLE DECIMAL	FLOAT NUMBER	FLOAT NUMERIC DECIMAL
Date and Time	DATE DATETIME TIMESTAMP	DATE TIMESTAMP TIMESTAMP WITH TIMEZONE INTERVAL	DATE TIME DATETIME DATETIMEOFFSET
Character	CHAR VARCHAR TEXT	CHAR VARCHAR2 LONG	CHAR VARCHAR TEXT
Binary	TINYBLOB MEDIUMBLOB LONGBLOB	BLOB BFILE RAW	BINARY VARBINARY IMAGE
Spatial	POINT POLYGON GEOMETRY	SDO_GEOMETRY	POINT POLYGON
Miscellaneous	ENUM BOOLEAN BIT	XMLTYPE	MONEY XML BIT

Category	Data type	Value
Integer	INT	-9281344
Decimal	FLOAT	3.1415
Character	CHAR	Chicago
Date and time	DATETIME	12/25/2020 10:35:00
Binary	BLOB	1001011101 . . .
Spatial	POINT	(2.5, 33.44)
Miscellaneous	MONEY	99.95 US Dollars

Example Data Types and Values

A Database Table

Student ID	Last Name	First Name	Email	Major
100	Skywalker	Luke	lskywalker@lewisu.edu	Farming
200	Skywalker	Anakin	askywalker@lewisu.edu	Sith
300	Palpatine	Rey	rpalpatine@lewisu.edu	Scavenger
400	Amidala	Padme	pamidala@lewisu.edu	Admin
500	Organa	Leia	lorgana@lewisu.edu	Admin
600	Calrissian	Lando	lcalrissian@lewisu.edu	Business
700	Solo	Han	hsolo@lewisu.edu	Business

Not Valid:
Multiple Entries per Cell

Student ID	Last Name	First Name	Email	Major
100	Skywalker	Luke	lskywalker@lewisu.edu	Farming, Education
200	Skywalker	Anakin	askywalker@lewisu.edu	Sith
300	Palpatine	Rey	rpalpatine@lewisu.edu	Scavenger
400	Amidala	Padme	pamidala@lewisu.edu	Admin
500	Organa	Leia	lorgana@lewisu.edu	Admin, Military
600	Calrissian	Lando	lcalrissian@lewisu.edu	Business
700	Solo	Han	hsolo@lewisu.edu	Business, Military

Not Valid:
Table with Required Row Order

Student ID	Last Name	First Name	Email	Major
100	Skywalker	Luke	lskywalker@lewisu.edu	Farming
				Education
200	Skywalker	Anakin	askywalker@lewisu.edu	Sith
300	Palpatine	Rey	rpalpatine@lewisu.edu	Scavenger
400	Amidala	Padme	pamidala@lewisu.edu	Admin
500	Organa	Leia	lorgana@lewisu.edu	Admin
				Military
600	Calrissian	Lando	lcalrissian@lewisu.edu	Business
700	Solo	Han	hsolo@lewisu.edu	Business
				Military

Null Values

Represents value for an attribute that is currently unknown or not applicable for row.

Deals with incomplete or exceptional data.

Represents the absence of a value and is not the same as zero or spaces, which are values.

Primary Keys and Foreign Keys

Primary Keys

- Column(s) used to identify a row
- A multiple columns keys is called a composite primary key
- Columns containing primary keys
  - All values must be unique
  - Cannot contain null values
- **Entity Integrity Constraint**: In a relation, no attribute of a primary key can be null.

Keys

- Superkey
  - An column, or set of columns, that uniquely identifies a row within a table.
- Candidate Key
  - Superkey (K) such that no proper subset is a superkey within the relation.
  - In each column of a tables, values of K uniquely identify that row (uniqueness).
  - No proper subset of K has the uniqueness property (irreducibility).
- Primary Key
  - Candidate key selected to identify tuples uniquely within relation.
- Alternate Keys
  - Candidate keys that are not selected to be primary key.

More Details About Keys

- Super key – can include a primary key plus alternate keys
- Primary key – one single column that uniquely identifies rows
- Candidate keys – columns that together uniquely identify rows
- Alternate keys – if more than one primary key exists, then choose one as primary and the rest are alternates
- Composite keys – when two or more keys are necessary to act as a primary key.

Keys?

Superkey?

Composite Key?

Primary Key?

Alternate Key?

firstname	lastname	birthdate	rating	teamid
Julie	Zion	2000-08-08	B	1
Britta	Zarinsky	2001-09-09	B	7
Jenna	Koester	2001-12-26	B	5
Melissa	Kimble	2000-05-31	C	7
Jessica	Anderson	2001-03-16	A	3
Leslie	Ericson	2000-12-21	B	3
Jessica	Goodman	2000-03-16	A	1
Marci	Barber	2001-05-15	A	4
Nancy	Dorman	2001-12-12	A	2
Julie	Fraser	2000-04-04	C	4
Kelly	James	2002-07-01	A	3
Sherrie	Green	2000-08-22	A	4
Priscilla	Pearson	2002-02-03	B	3
Kerri	Stone	2000-04-04	C	5
Katie	Remmen	2002-04-18	A	1
Katherine	Akong	2000-02-02	C	NULL
Lilly	Taboas	2001-05-21	B	NULL
Jessica	Benjamin	2001-10-31	A	2

No primary key. Composite keys are lastname plus firstname plus birthdate. An alternate key could include the teamid. All of these columns together are the super key.

Keys?

What is different here? How does it affect the other keys?

id	firstname	lastname	birthdate	rating	teamid
2	Julie	Zion	2000-08-08	B	1
3	Britta	Zarinsky	2001-09-09	B	7
4	Jenna	Koester	2001-12-26	B	5
5	Melissa	Kimble	2000-05-31	C	7
6	Jessica	Anderson	2001-03-16	A	3
7	Leslie	Ericson	2000-12-21	B	3
8	Jessica	Goodman	2000-03-16	A	1
9	Marci	Barber	2001-05-15	A	4
10	Nancy	Dorman	2001-12-12	A	2
11	Julie	Fraser	2000-04-04	C	4
12	Kelly	James	2002-07-01	A	3
13	Sherrie	Green	2000-08-22	A	4
14	Priscilla	Pearson	2002-02-03	B	3
15	Kerri	Stone	2000-04-04	C	5
16	Katie	Remmen	2002-04-18	A	1
17	Katherine	Akong	2000-02-02	C	NULL
18	Lilly	Taboas	2001-05-21	B	NULL
19	Jessica	Benjamin	2001-10-31	A	2

Foreign Key

- Column, or set of column, within one table that matches a primary key of some (possibly same) table.
- Foreign keys do not obey the same rules as primary keys:
  - Foreign key values may be repeated.
  - Foreign key values may be NULL
  - Non-NULL foreign key values must match some primary key value.

Foreign Key Example

id	firstname	lastname	birthdate	rating	teamid
2	Julie	Zion	2000-08-08	B	1
3	Britta	Zarinsky	2001-09-09	B	7
4	Jenna	Koester	2001-12-26	B	5
5	Melissa	Kimble	2000-05-31	C	7
6	Jessica	Anderson	2001-03-16	A	3
7	Leslie	Ericson	2000-12-21	B	3
8	Jessica	Goodman	2000-03-16	A	1
9	Marci	Barber	2001-05-15	A	4
10	Nancy	Dorman	2001-12-12	A	2
11	Julie	Fraser	2000-04-04	C	4
12	Kelly	James	2002-07-01	A	3
13	Sherrie	Green	2000-08-22	A	4
14	Priscilla	Pearson	2002-02-03	B	3
15	Kerri	Stone	2000-04-04	C	5
16	Katie	Remmen	2002-04-18	A	1
17	Katherine	Akong	2000-02-02	C	NULL
18	Lilly	Taboas	2001-05-21	B	NULL
19	Jessica	Benjamin	2001-10-31	A	2

players

teamid is a foreign key
teamid references the id column
in teams

id	name	color
1	Rockets	Red/White
2	Comets	Blue/White
3	Bulldogs	Red/Black
4	Tar Heels	Orange/Green
5	Tornadoes	Blue/Gold
6	Eagles	Orange/Blue
7	Blue Devils	Maroon/Gold

teams