



Διάλεξη 06: Σχεσιακό Μοντέλο I (Relational Data Model)

Στην ενότητα αυτή θα μελετηθούν τα εξής επιμέρους θέματα:
Εισαγωγή στις έννοιες:

- Εισαγωγή στο Σχεσιακό Μοντέλο
- Ορισμοί Σχεσιακού Μοντέλου
- Σχεσιακοί Κανόνες Ακεραιότητας

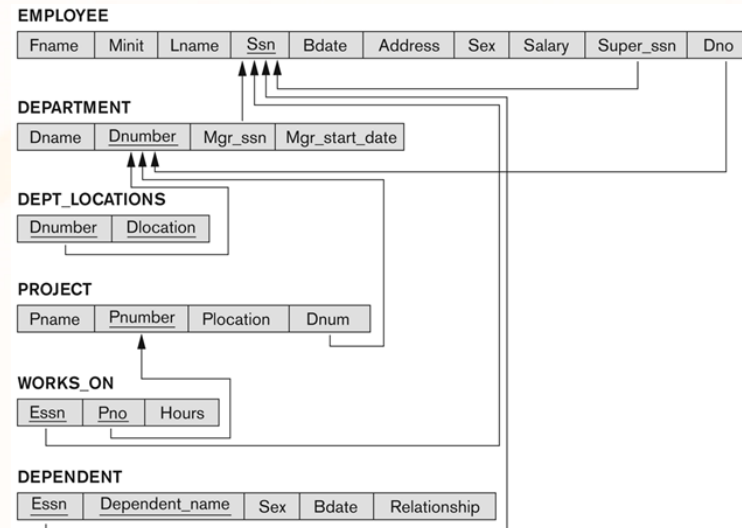
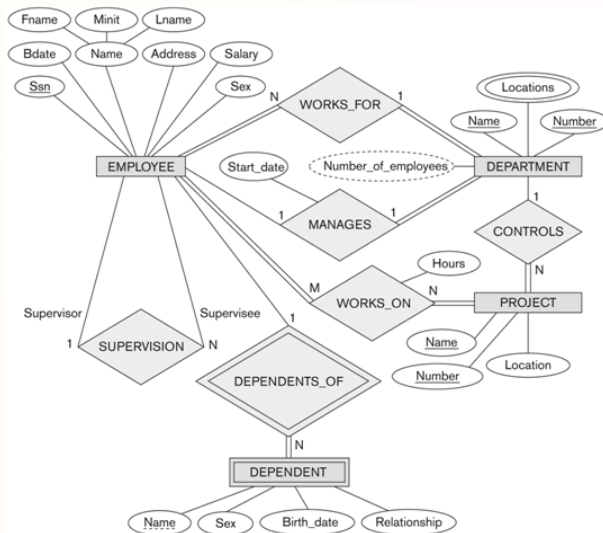
Διδάσκων: Παναγιώτης Ανδρέου

Εισαγωγή στο Σχεσιακό Μοντέλο

- Το **Μοντέλο Οντοτήτων-Συσχετίσεων (ER Model)**, μέσα από τις άτυπες έννοιες του (π.χ., Οντότητες, Συσχετίσεις) γίνεται **εύκολα αντιληπτό από τους σχεδιαστές και πελάτες**
- Σε αυτή την ενότητα θα δούμε το **Σχεσιακό Μοντέλο (Relational Model)**, το οποίο κάνει χρήση του μαθηματικού υποβάθρου των Σχέσεων για να εκφράσει τη δομή μίας βάσης δεδομένων.
- Το **Σχεσιακό Μοντέλο** ορίστηκε το 1970 από τον Βρετανό ερευνητή *Edgar F. Codd* στο *IBM Research*
 - "A Relational Model of Data for Large Shared Data Banks". "Codd, E.F.(1970), *Communications of the ACM* 13(6):377-387
 - Σε αυτό το μοντέλο, τα δεδομένα αναπαριστώνται σε **πίνακες** πάνω στους οποίους ορίζονται **διάφοροι περιορισμοί**.
 - Το μοντέλο προκάλεσε μια **επανάσταση** στο χώρο των βάσεων δεδομένων λόγω της **απλότητας** και του **μαθηματικού του υπόβαθρου**
 - **Ιστορικό: 1969:** Το Σχεσιακό Μοντέλο υλοποιείται από τη βάση **IBM System R**, **1970:** Η IBM δημιουργεί την **SEQUEL** (προπομπό της **SQL**), **1981:** Ο **Codd** παίρνει το **Turing Award** στη πληροφορική, **1985:** Η **IBM** κάνει την **SQL** Πατέντα (US Pat. 4,506,326)., **Σήμερα:** Το Σχεσιακό Μοντέλο υλοποιείται από τις περισσότερες σύγχρονες βάσεις δεδομένων (*Oracle, IBM DB2, SQL Server, PostgreSQL, MySQL*).

Εισαγωγή στο Σχεσιακό Μοντέλο (συν.)

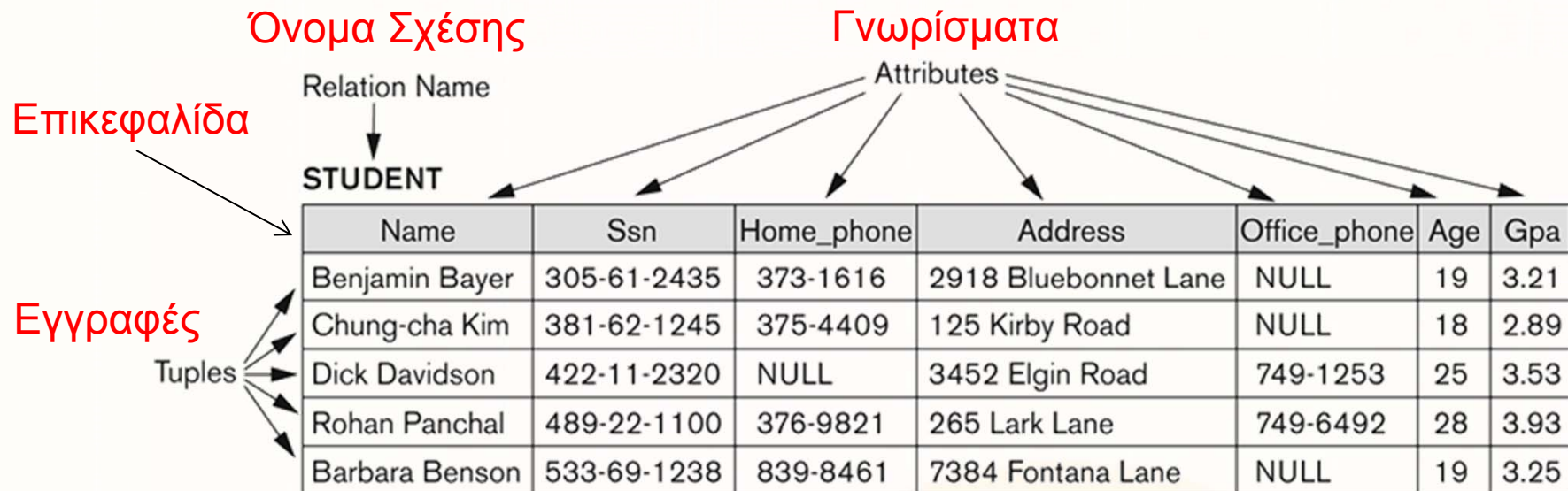
- **Στόχος του Σχεσιακού Μοντέλου:** Να προσφέρει μια δηλωτική (*declarative*) μέθοδο για τον ορισμό **δεδομένων** και **επερωτήσεων**
- Πάνω στο σχεσιακό μοντέλο στηρίζεται σήμερα η **Σχεσιακή Άλγεβρα (*Relational Algebra*)** (και κατ' επέκταση η *SQL*)
 - Εάν και υπάρχουν κάποιες **λεπτές διαφορές** μεταξύ του **Σχεσιακού Μοντέλου** και του πως **αναπαριστά** τα δεδομένα μια βάση δεδομένων.
- Το σχεσιακό μοντέλο είναι **πιο κοντά στο φυσικό σχήμα** τις βάσης δεδομένων
- Δεν μπορεί να εκφράσει όλους τους περιορισμούς ενός *ER* μοντέλου
→ χρησιμοποιείται σε συνδυασμό



Σχεσιακό Μοντέλο: Άτυποι Ορισμοί

- Στο **Σχεσιακό Μοντέλο (Relational Model)** κάνει χρήση της μαθηματικής έννοιας της **Σχέσης (Relation)**
- Τυπικά, μια **Σχέση r** είναι ένα σύνολο **πλειάδων (tuples)** $r = \{t_1, t_2, \dots, t_M\}$, όπου κάθε πλειάδα είναι μια **διατεταγμένη λίστα τιμών $t = \langle v_1, v_2, \dots, v_N \rangle$** .
- Συνεπώς, μια σχέση είναι ουσιαστικά ένας **πίνακας $M \times N$ (M γραμμών και N στηλών)**!
- Κάθε γραμμή αναπαριστά συγκεκριμένα γεγονότα τα οποία αναφέρονται σε κάποια οντότητα (**entity**) ή συσχέτιση (**relationship**) του πραγματικού κόσμου.
 - Κάθε **γραμμή row** ονομάζεται και **πλειάδα (tuple)** ή **εγγραφή (record)**
 - Κάθε **στήλη (column)** ονομάζεται και **χαρακτηριστικό (characteristic)** ή **γνώρισμα (attribute)**
 - Η επικεφαλίδα κάθε στήλης προσδιορίζει το νόημα της εν λόγω στήλης

Σχεσιακό Μοντέλο: Παράδειγμα



Τυπικά, η **επικεφαλίδα** ΔΕΝ είναι μέρος της **σχέσης** αλλά περιλαμβάνεται στις **Μέτα-πληροφορίες** της.

Που αποθηκεύονται οι τέτοιες Μέτα-πληροφορίες σε με σχεσιακή βάση δεδομένων;

Σχεσιακό Μοντέλο: Άτυποι Ορισμοί

- **Κλειδί (Key) μιας Σχέσης:** Η ανάθεση τιμής ενός ή περισσότερων γνωρισμάτων τα οποία προσδιορίζουν μοναδικά την γραμμή στον πίνακα.
 - Ο ορισμός είναι αντίστοιχος με αυτόν που δώσαμε στο πλαίσιο του ER όπου ορίσαμε ως το **Κλειδί μιας Οντότητας (Key)** το πεδίο(α) που αναγνωρίζουν μοναδικά μια Οντότητα
 - Στο **Σχεσιακό Μοντέλο** μια **σχέση** είναι είτε μια **Οντότητα** ή μια **Συσχέτιση** του ER Μοντέλου
- Στο παράδειγμα του πίνακα STUDENT, το SSN είναι προφανώς το κλειδί.
- Κάποτε η θέση της γραμμής (row-id) ή σειριακοί αριθμοί χρησιμοποιούνται ως κλειδιά
 - Ονομάζονται **Τεχνητά Κλειδιά (Artificial keys)** ή **Υποκατάστατα Κλειδιά (surrogate keys)**

Σχήμα Σχέσης

- Το **Σχήμα (Schema)** ή περιγραφή μιας **Σχέσης**:
 - Θα αναφέρεται ως **$R(A_1, A_2, \dots, A_n)$**
 - **R** είναι το **όνομα** της σχέσης
 - Τα **γνωρίσματα** της σχέσης είναι τα **A_1, A_2, \dots, A_n**
- **Παράδειγμα:**
CUSTOMER (Cust-id, Cust-name, Address, Phone#)
 - CUSTOMER δηλώνει το **όνομα** της σχέσης
 - Η σχέση ορίζεται από τα **γνωρίσματα**: Cust-id, Cust-name, Address, Phone#
- Κάθε γνώρισμα ορίζεται στο **πεδίο ορισμού (domain)**
 - π.χ., το domain του Cust-id είναι 6 ψηφία (ακολουθεί ορισμός)
- Ο **βαθμός (degree ή arity)** μιας σχέσης είναι το πλήθος **n** των γνωρισμάτων του σχήματος.

n-άδα ή Πλειάδα

- Μια **πλειάδα (tuple)** είναι ένα **διατεταγμένο*** σύνολο τιμών v_1, v_2, \dots, v_n το οποίο εμπερικλείεται στον ακόλουθο συμβολισμό '< ... >'
 - Κάθε τιμή v_j ($1 \leq j \leq n$) της πλειάδας **ορίζεται** στο αντίστοιχο **πεδίο ορισμού** του συγκεκριμένου γνωρίσματος A_j ($1 \leq j \leq n$) ή είναι **κενό**.
 - * εάν και θα χαλαρώσουμε αυτό τον περιορισμό αργότερα
- **Παράδειγμα:** CUSTOMER (Cust-id, Cust-name, Address, Phone#)
- Μια γραμμή της σχέσης CUSTOMER είναι μια 4-άδα η οποία αποτελείται από 4 τιμές, για παράδειγμα:
 - <632895, "John Smith", "101 Main St. Atlanta, GA 30332", "(404) 894-2000">
- Μια σχέση είναι ένα **σύνολο (set)** τέτοιων πλειάδων.

Πεδίο Ορισμού

- Το **Πεδίο Ορισμού (Domain – dom(Ai))** είναι το σύνολο των δυνατών **ατομικών** τιμών ενός γνωρίσματος (δηλ., δεν κάνει νόημα να διασπαστούν οι τιμές περαιτέρω), π.χ.,
 - **Int** → -2^{31} (-2,147,483,648) μέχρι $2^{31}-1$ (2,147,483,647)
 - **Τηλεφωνικοί Αριθμοί Πελατών** → Το σύνολο όλων των έγκυρων αριθμών τηλεφώνων.
- Το Πεδίου Ορισμού έχει ένα **Λογικό Ορισμό (εξήγηση)**:
 - Π.χ.: “nicosia_phone_numbers”: *το σύνολο των αριθμών που είναι έξι ψηφία και ξεκινούν με 22.*
- Επιπλέον, έχει **Όνομα, Τύπο και Μορφοποίηση**:
 - **Τύπος Δεδομένων nicosia_phone_numbers** : Συμβολοσειρά
 - “22-892700”
 - **Μορφοποίηση nicosia_phone_numbers** : Οι αριθμοί πρέπει να έχουν την μορφή 22-dddddd, όπου d είναι ένα δεκαδικό ψηφίο.
- Τέλος, ίσως να απαιτείται και επιπλέον πληροφορία (π.χ., μονάδα μέτρησης σε ένα πεδίο **persons_weight: kg, g**)

Κατάσταση/Στιγμιότυπο Σχέσης

- Η Κατάσταση ή Στιγμιότυπο Σχέσης (Relation State – $r(R)$) είναι ένα υποσύνολο (ακόμη και κενό) του Καρτεσιανού Γινομένου των πεδίων ορισμού των γνωρισμάτων
 - Συνεπώς, το $r(R)$ είναι ένα σύνολο πλειάδων (tuples) $r = \{t_1, t_2, \dots, t_M\}$, όπου κάθε πλειάδα είναι μια διατεταγμένη ακολουθία τιμών $t = \langle v_1, v_2, \dots, v_N \rangle$.
 - Κάθε τιμή v_j είναι ένα στοιχείο του $\text{dom}(A_j)$ ή NULL
- Το $r(R)$ ονομάζεται συχνά και έκταση σχέσης (relation extension) και το σχήμα R μιας σχέσης ως πρόθεση σχέσης (relation intension)
- Πληθυκός Αριθμός Σχέσης (Cardinality) ($|r(R)|$): Ο αριθμός των πλειάδων μιας σχέσης.
 - δηλ., ο αριθμός γραμμών του πίνακα της σχέσης

Σύνοψη Ορισμών Σχεσιακού Μοντέλου

Στιγμιότυπο $r(\text{STUDENT})$

Όνομα Σχέσης

Γνωρίσματα

Πλειάδες

Name	Ssn	Home_phone	Address	Office_phone	Age	Gpa
Benjamin Bayer	305-61-2435	373-1616	2918 Bluebonnet Lane	NULL	19	3.21
Chung-cha Kim	381-62-1245	375-4409	125 Kirby Road	NULL	18	2.89
Dick Davidson	422-11-2320	NULL	3452 Elgin Road	749-1253	25	3.53
Rohan Panchal	489-22-1100	376-9821	265 Lark Lane	749-6492	28	3.93
Barbara Benson	533-69-1238	839-8461	7384 Fontana Lane	NULL	19	3.25

$r(\text{STUDENT})$

Βαθμός (Degree) = $|R| = 7$ (γνωρίσματα)

Πληθικός Αριθμός (Cardinality) = $|r(\text{STUDENT})| = 5$ (πλειάδες)

Σύνοψη Ορισμών Σχισιακού Μοντέλου

- $R(A_1, A_2, \dots, A_n)$ είναι το **σχήμα** της σχέσης.
- R είναι το **όνομα** της σχέσης
- A_1, A_2, \dots, A_n είναι τα **γνωρίσματα** της σχέσης
 - $\text{dom}(A_j)$: Το πεδίο ορισμού του A_j
- $r(R)$: είναι μια συγκεκριμένη **κατάσταση (state)** της σχέσης R
 - Αυτό είναι ένα σύνολο από πλειάδες (έγγραφες)
 - $r(R) = \{t_1, t_2, \dots, t_m\}$ όπου κάθε t_i είναι μια n -άδα
 - $t_i = \langle v_1, v_2, \dots, v_n \rangle$ όπου κάθε τιμή v_j ανήκει στο $\text{dom}(A_j)$
 - Πληθυκός Αριθμός Σχέσης: $|r(R)| = m$
 - Βαθμός Σχέσης: $|R| = n$

Σύνοψη Ορισμών Σχισιακού Μοντέλου

- Έστω ότι $R(A_1, A_2)$ είναι ένα σχήμα σχέσης:
 - Υποθέστε ότι $\text{dom}(A_1) = \{0,1\}$
 - Υποθέστε επίσης ότι $\text{dom}(A_2) = \{a,b,c\}$
- Τότε: $\text{dom}(A_1) \times \text{dom}(A_2)$ είναι όλοι δυνατοί συνδυασμοί:
 $\{ \langle 0,a \rangle, \langle 0,b \rangle, \langle 0,c \rangle, \langle 1,a \rangle, \langle 1,b \rangle, \langle 1,c \rangle \}$
- Η κατάσταση της σχέσης είναι υποσύνολο του καρτεσιανού γινομένου των πεδίων ορισμού $r(R) \subseteq \text{dom}(A_1) \times \text{dom}(A_2)$
 - Π.χ., $r(R)$ μπορεί να είναι $\{\}$ ή $\{\langle 0,b \rangle\}$ ή $\{\langle 0,a \rangle, \langle 0,b \rangle, \langle 1,c \rangle\}$ ή οποιοδήποτε άλλο στοιχείο του Δυναμοσυνόλου (PowerSet) του $\text{dom}(A_1) \times \text{dom}(A_2)$
 - Το **Δυναμοσύνολο (powerset) $P(A)$** ενός σύνολο A είναι το σύνολο όλων των υποσυνόλων του A .
 - Αν $A = \{a,b\}$ τότε $P(A) = \{ \{\}, \{a\}, \{b\}, \{a,b\} \}$
 - Ο **πληθυκός αριθμός του δυναμοσυνόλου $|P(A)| = 2^{|A|}$**
 - Συνεπώς υπάρχουν $2^6 = 64$ διαφορετικά στιγμιότυπα του $r(R)$

Σύνοψη Ορισμών Σχεσιακού Μοντέλου

- Πιο κάτω φαίνεται η αντιστοίχιση της άτυπης ορολογίας προς την επίσημη ορολογία

<u>Άτυποι Όροι</u>	<u>Επίσημοι Όροι</u>
Πίνακας	Σχέση (Relation)
(Κεφαλίδα) Στήλης	Γνώρισμα (Attribute)
Όλες οι δυνατές τιμές μιας στήλης	Πεδίο Ορισμού (Domain)
Γραμμή	Πλειάδα (Tuple)
Δήλωση Πίνακα	Σχήμα Σχέσης (Schema)
Δεδομένα Πίνακα	Στιγμιότυπο Σχέσης (State)

Χαρακτηριστικά των Σχέσεων

1. Διάταξη των πλειάδων σε μια σχέση $r(R)$:

- Οι πλειάδες ΔΕΝ είναι **διατεταγμένες** ή ταξινομημένες, παρόλο ίσως να παρουσιάζονται έτσι.
- Μια σχέση είναι ουσιαστικά ένα **σύνολο πλειάδων**.
- Τα σύνολα εξ' ορισμού δεν είναι διατεταγμένα.
- **Ορισμός Συνόλου (Cantor 1895)**: Μια συλλογή ή ομάδα αντικειμένων ή στοιχείων ή μελών
- Η **διάταξη** ΔΕΝ αποτελεί μέρος του ορισμού μιας **σχέσης** αφού μια σχέση προσπαθεί να αναπαραστήσει κάποια γεγονότα σε **λογικό επίπεδο**.

Διάταξη Πλειάδων

STUDENT

Name	Ssn	Home_phone	Address	Office_phone	Age	Gpa
Benjamin Bayer	305-61-2435	373-1616	2918 Bluebonnet Lane	NULL	19	3.21
Chung-cha Kim	381-62-1245	375-4409	125 Kirby Road	NULL	18	2.89
Dick Davidson	422-11-2320	NULL	3452 Elgin Road	749-1253	25	3.53
Rohan Panchal	489-22-1100	376-9821	265 Lark Lane	749-6492	28	3.93
Barbara Benson	533-69-1238	839-8461	7384 Fontana Lane	NULL	19	3.25

$$r(\text{Student1}) = r(\text{Student2})$$

STUDENT2

Name	Ssn	Home_phone	Address	Office_phone	Age	Gpa
Dick Davidson	422-11-2320	NULL	3452 Elgin Road	749-1253	25	3.53
Barbara Benson	533-69-1238	839-8461	7384 Fontana Lane	NULL	19	3.25
Rohan Panchal	489-22-1100	376-9821	265 Lark Lane	749-6492	28	3.93
Chung-cha Kim	381-62-1245	375-4409	125 Kirby Road	NULL	18	2.89
Benjamin Bayer	305-61-2435	373-1616	2918 Bluebonnet Lane	NULL	19	3.21

Χαρακτηριστικά των Σχέσεων (συν.)

2. Διάταξη Γνωρισμάτων σε μια Πλειάδα:

- Ο ορισμός της πλειάδας που χρησιμοποιήσαμε θεωρεί ότι τα γνωρίσματα είναι διατεταγμένα.
 - «Πλειάδα είναι ένα **διατεταγμένο σύνολο** τιμών v_1, v_2, \dots, v_m ...»)
- Τώρα θα **χαλαρώσουμε** αυτό τον περιορισμό αφού σε λογικό επίπεδο η **διάταξη των πλειάδων** δεν παίζει **σημαντικό** ρόλο, π.χ.,
 - Student (Name, SSN, Home_Phone, Address,...)
 - Student (SSN, Name, Home_Phone, Address,...)
- Συνεπώς, θα θεωρούμε ότι μια **πλειάδα** είναι ένα σύνολο από ζεύγη της μορφής (<γνώρισμα>, <τιμή>) όπου κάθε τιμή ορίζεται στο **dom(γνώρισμα)**.

Χαρακτηριστικά των Σχέσεων

3. Τιμές και NULL Τιμές στις Πλειάδες:

- Κάθε τιμή σε μια πλειάδα είναι μια ατομική τιμή η οποία ορίζεται στο Πεδίο Ορισμού του εν λόγω γνωρίσματος
- Κάθε Πλειότιμη Τιμή (**Multi-valued Attribute**) αναπαριστάται από μια επί μέρους σχέση
 - Π.χ., Γνώρισμα «Locations» σε μια σχέση Department
- Κάθε Σύνθετη Τιμή (**Composite attribute**) αναπαριστάται από τα απλά συστατικά τους.
 - Π.χ., Γνώρισμα «Name» αναπαριστάται στη σχέση από “Fname”, “Minit” και “Lname”
- **NULL Τιμές: Κενές Τιμές** (δεν ισχύουν ή δεν είναι γνωστές)
 - Π.χ., δεν καταχωρήσαμε το Office_phone κάποιου Employee
- Τιμές που μπορεί να **μην ισχύουν** για κάποια πλειάδα
 - Π.χ., Ένας Υπάλληλος δεν δουλεύει σε γραφείο αλλά σε κάποιο άλλο χώρο (και δεν έχει γραφείο)

NULL Τιμές

- Γιατί μας απασχολούν οι NULL τιμές;
 - Γιατί **σπαταλούν χώρο**.
 - Γιατί η **σύγκριση** τους μπορεί να οδηγήσει σε **ασάφειες**.
 - Π.χ., Βρες όλους τους υπαλλήλους που έχουν το ίδιο τηλέφωνο.
 - Οι σχεσιακές βάσεις υλοποιούν μια τριαδική λογική: **{TRUE, FALSE, UNKNOWN}** αντί δυαδικής **{TRUE, FALSE}** στους τελεστές σύγκρισης για να λύσουν αυτό το πρόβλημα,
- Κατά τον σχεδιασμό βάσεων προσπαθούμε να **αποφεύγουμε** τις NULL τιμές όσο το δυνατό περισσότερο.

STUDENT

Name	Ssn	Home_phone	Address	Office_phone	Age	Gpa
Dick Davidson	422-11-2320	NULL	3452 Elgin Road	749-1253	25	3.53
Barbara Benson	533-69-1238	839-8461	7384 Fontana Lane	NULL	19	3.25
Rohan Panchal	489-22-1100	376-9821	265 Lark Lane	749-6492	28	3.93
Chung-cha Kim	381-62-1245	375-4409	125 Kirby Road	NULL	18	2.89
Benjamin Bayer	305-61-2435	373-1616	2918 Bluebonnet Lane	NULL	19	3.21

Σχεσιακοί Κανόνες Ακεραιότητας

- Μέχρι τώρα είδαμε διάφορους **ορισμούς** και τα **χαρακτηριστικά** τα οποία αφορούν μια **Σχέση**.
- Σε μια σχεσιακή βάση ωστόσο υπάρχουν **πολλές Σχέσεις**
 - Θυμηθείτε τις **οντότητες** και τις **συσχετίσεις** του **ER Μοντέλου**, οι οποίες αναπαριστώνται και οι δυο σαν Σχέσεις (Πίνακες) στο Σχεσιακό Μοντέλο.
- Σε μια σχεσιακή βάση υπάρχουν επίσης **πολλοί τύποι περιορισμών (Relational Integrity Constraints)**, οι οποίοι αναφέρονται
 - Σε μια **σχέση** (π.χ., περιορισμός κλειδιού, πεδίου ορισμού και οντότητας)
 - Σε **πολλαπλές σχέσεις** (π.χ., περιορισμός αναφορικής ακεραιότητας)

Σχεσιακοί Κανόνες Ακεραιότητας (συν.)

- **Σχεσιακοί Περιορισμοί (Relational Constraints)** είναι συνθήκες οι οποίες πρέπει να ισχύουν για κάθε **έγκυρη κατάσταση σχέσης (state)**.
- Στο σχεσιακό μοντέλο υπάρχουν τρεις τύποι **περιορισμών**:
 - A. Κλειδιού (Key constraints)**
 - **Άτυπα:** Κάθε σχέση έχει ένα πρωτεύων κλειδί.
 - B. Ακεραιότητας Οντοτήτων (Entity integrity constraints)**
 - **Άτυπα:** Το πρωτεύων κλειδί δεν μπορεί να είναι NULL
 - C. Αναφορικής Ακεραιότητας (Referential integrity constraints)**
 - **Άτυπα:** Εάν μια πλειάδα A αναφέρεται σε άλλη πλειάδα B τότε η B πρέπει να υπάρχει.
- Τέλος, υπάρχει και ο **Περιορισμός Πεδίου Ορισμού (Domain Constraint)** ο οποίος εξυπακούεται.
 - Η τιμή κάθε πλειάδας πρέπει να ορίζεται στο **αντίστοιχο της πεδίο ορισμού** (ή μπορεί να είναι **NULL**, εάν επιτρέπεται για το εν λόγω γνώρισμα)

Σχισιακοί Κανόνες Ακεραιότητας - Συζήτηση

