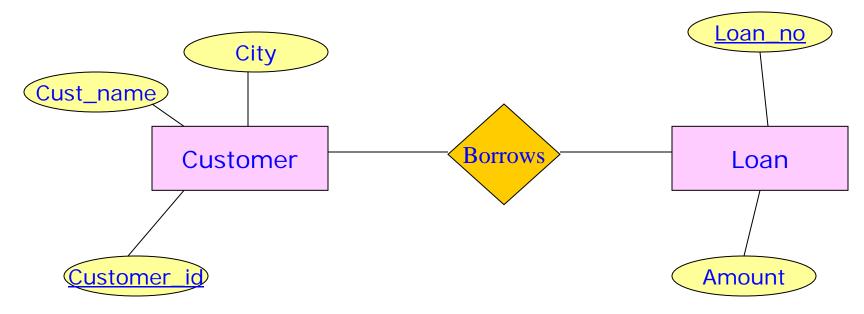
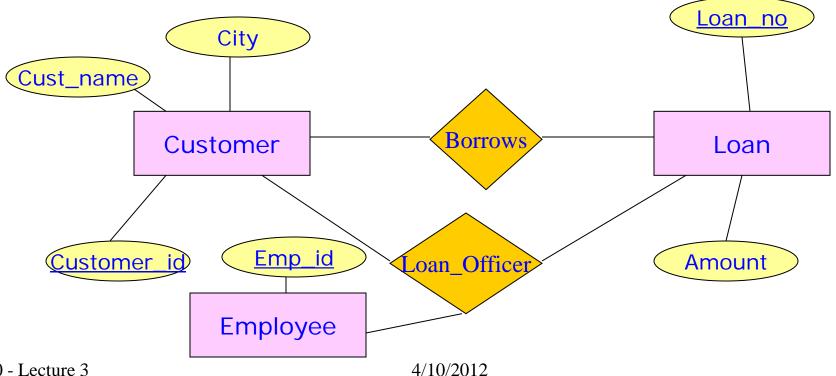
- Συνάθροιση (Aggregation)
- Ένας από τους περιορισμούς του μοντέλου οντοτήτων-σχέσεων είναι ότι δεν είναι δυνατός ο ορισμός σχέσεων μεταξύ σχέσεων
- Τέτοιες σχέσεις είναι συχνά απαραίτητες.
- Για παράδειγμα:

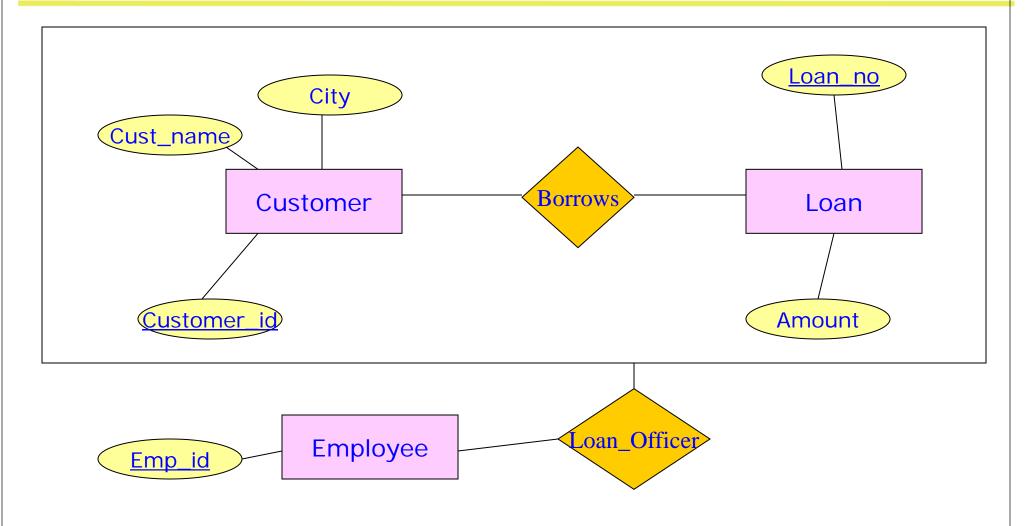


- Υποθέστε ότι σε κάθε ζεύγος πελάτη-δανείου αντιστοιχίζεται ένας υπάλληλος ως υπεύθυνος. Πως θα αναπαρασταθεί αυτή η σχέση?
- Ενας τρόπος θα ήταν να εισαχθεί μια καινούργια σχέση μεταξύ πελατών και δανείων.



- Μπορούν οι σχέσεις Borrows και Loan\_Officer να συνδυαστούν σε μία σχέση?
  - Αν συνδυαστούν, αυτό σημαίνει ότι ένας υπάλληλος πρέπει να αντιστοιχηθεί σε κάθε ζεύγος πελάτη-δανείου.
  - Υπάρχει επίσης πλεονάζουσα πληροφορία: κάθε ζεύγος πελάτηδανείου στη σχέση Loan\_Officer ανήκει και στη σχέση Borrows.
  - Εναλλακτικά, μπορεί να εισαχθεί το γνώρισμα loan\_officer στη σχέση Borrows.
  - Πως όμως θα βρούμε τα ζεύγη πελατών δανείων για τα οποία ένας υπάλληλος είναι υπεύθυνος?

- Το πρόβλημα μπορεί να επιλυθεί με τη χρήση του μηχανισμού της συνάθροισης, δηλαδή να θεωρήσουμε ότι η σχέση Borrows και οι σχετιζόμενες οντότητες αποτελούν μια οντότητα υψηλότερου επιπέδου.
- Μια τέτοια οντότητα συμπεριφέρεται όπως κάθε άλλη οντότητα. Ποιο θα είναι το αναγνωριστικό της?
  - Το αναγνωριστικό θα σχηματιστεί από τα αναγνωριστικά των οντοτήτων που συμμετέχουν στη σχέση Borrows
- Συμβολισμός: χρησιμοποιείται ένα παραλληλόγραμμο το οποίο περικλείει τις οντότητες και τη σχέση



- Συμβολισμός για την αναπαράσταση δεδομένων: πίνακες
- Κανόνες για τον καθορισμό της σωστής δομής των πινάκων
- Τελεστές για το χειρισμό πινάκων: σχεσιακή άλγεβρα (relational algebra)
- Η σχεσιακή άλγεβρα χρησιμοποιείται και ως γλώσσα επερωτήσεων (query language)
- SQL (structured query language): πρότυπο γλωσσών επερωτήσεων σχεσιακών βάσεων δεδομένων, βασισμένη στη σχεσιακή άλγεβρα

- Παράδειγμα: CAP database
  - ΒΔ που χρησιμοποιείται από εταιρία χονδρικής ώστε να κρατά αρχείο των πελατών (customers), των προϊόντων (products) που πουλάει και των πρακτόρων (agents) που κάνουν παραγγελίες (orders) προϊόντων εκ μέρους των πελατών.
  - Οι πελάτες είναι εταιρίες λιανικής που παραγγέλνουν μεγάλες ποσότητες προϊόντων από την εταιρία χονδρικής.
  - Οι πελάτες, τα προϊόντα, οι πράκτορες και οι παραγγελίες χαρακτηρίζονται από μοναδικά προσδιοριστικά (identifiers).

- Ορολογία:

  - γραμμές / πλειάδες (rows / tuples)
  - στήλες / γνωρίσματα (columns / attributes)
  - Μια βάση δεδομένων είναι ένα σύνολο από σχέσεις με μοναδικά ονόματα
    - Π.χ., η βάση δεδομένων CAP ορίζεται ως CAP = {Customers, Agents, Products, Orders}
  - το σχήμα ή επικεφαλίδα (heading) ενός πίνακα είναι το σύνολο των ονομάτων των γνωρισμάτων της σχέσης που αναπαριστά
    - $\Pi.\chi$ . Head(Customers) = {cid, cname, city, discnt}

- Το σύνολο των σχημάτων όλων των σχέσεων ονομάζεται σχήμα της βάσης δεδομένων (database schema).
- Το σύνολο όλων των πλειάδων των σχέσεων λέγεται περιεχόμενο της βάσης δεδομένων (database content, extension). Για μια σχέση R, το περιεχόμενο συμβολίζεται ως cont(R)
- Κάθε γνώρισμα μιας σχέσης έχει έναν τύπο (type). Τα γνωρίσματα παίρνουν τιμές από ένα πεδίο τιμών (domain). Το σύνολο αυτό περιέχει τις σταθερές οι οποίες μπορούν να εμφανίζονται ως τιμές ενός γνωρίσματος.
- Το πεδίο τιμών ενός γνωρίσματος A συμβολίζεται ως domain(A). Μια σχέση R με σχήμα  $\{A_1, A_2, ..., A_m\}$  είναι ένα υποσύνολο του  $domain(A_1) \times domain(A_2) \times ... \times domain(A_m)$

- <u>Παράδειγμα</u>: CAP Database
  - head(Customers)={cid, cname, city, discnt}
  - $(c001, TipTop, Paris, 10.00) \in cont(Customers)$
  - **■** (c002, ACME, Kyoto, 8.00) ∈ domain(cid) × domain(cname) × domain(city) × domain(discnt)
  - **(**c002, ACME, Kyoto, 8.00) ∉ cont(Customers)

- Σχεσιακοί Κανόνες:
  - Ορίζουν επιλογές και περιορισμούς στη δομή των σχέσεων.
  - Έχουν σημαντικές επιπτώσεις στην «καλή συμπεριφορά» των δομών που χρησιμοποιούνται.
  - Υποδεικνύουν σημεία όπου είναι δυνατή η τυποποίηση.
  - Δεν ακολουθούνται πάντα από εμπορικά συστήματα.
- Κανόνας 1 (Πρώτη Κανονική Μορφή)
  - Μια σχέση είναι σε Πρώτη Κανονική Μορφή (1NF) αν δεν έχει πλειότιμα γνωρίσματα

- Κανόνας 1 (Πρώτη Κανονική Μορφή)
  - Παράδειγμα: η παρακάτω σχέση (Employees) δεν είναι σε 1NF

eid	ename	position	dependents
e001	Smith, John	Agent	Michael J.
			Susan R.
e002	Andrews, David	Superintendent	David M. Jr.
e003	Jones, Franklin	Agent	Andrew K.
			Mark W.
			Lois M.

- Κανόνας 1 (Πρώτη Κανονική Μορφή)
  - Πρόβλημα: πως θα αναπαρασταθεί πληροφορία σχετικά με τα εξαρτόμενα πρόσωπα σε σχέσεις σε 1NF?
  - 1η επιλογή: να αλλαχθεί το σχήμα ώστε να περιλάβει τόσα
     πρόσθετα γνωρίσματα όσα ο μέγιστος αριθμός εξαρτ. προσώπων

eid	ename	position	dep1	dep2	dep3
e001	Smith, John	Agent	Michael J.	Susan R.	
e002	Andrews, David	Superintendent	David M. Jr.		
e003	Jones, Franklin	Agent	Andrew K.	Mark W.	Lois M.

Μη αποδοτική λύση: σπατάλη χώρου, δυσκολία στις ερωτήσεις

- Κανόνας 1 (Πρώτη Κανονική Μορφή)
  - 2η επιλογή: να δημιουργηθεί μια νέα σχέση για τους υπαλλήλους
     και τα εξαρτόμενα από αυτούς πρόσωπα

eid	ename	position
e001	Smith, John	Agent
e002	Andrews, David	Superintendent
e003	Jones, Franklin	Agent

eid	dependent	
e001	Michael J.	
e001	Susan R.	
e002	David M. Jr.	
e003	Andrew K.	
e003	Mark W.	
e003	Lois M.	

Προτιμότερη λύση: η περιττή πληροφορία είναι η ελάχιστη δυνατή.

15

- Υπάρχουν συστήματα (extended RDBMS) τα οποία επιτρέπουν τον ορισμό σχέσεων που δεν είναι σε 1NF.
- Κανόνας 2: Πρόσβαση βάσει περιεχομένου (row access by content)
  - οι πλειάδες μιας σχέσης συνιστούν ένα μη-διατεταγμένο σύνολο. Η πρόσβαση στις πλειάδες γίνεται μόνο βάσει του περιεχομένου τους
  - ερωτήσεις της μορφής «ποια είναι η 3η πλειάδα της σχέσης?» είναι μη-έγκυρες
  - οι ερωτήσεις πρέπει να βασίζονται στις τιμές των γνωρισμάτων
  - πολλά εμπορικά συστήματα παραβιάζουν τον κανόνα αυτό και παρέχουν ένα εσωτερικό αναγνωριστικό πλειάδων (tuple id). Πρόσβαση βάσει του tuple id ενδέχεται να έχει πλεονεκτήματα από την άποψη της απόδοσης του συστήματος.

- Κανόνας 3: Μοναδικότητα πλειάδων (unique rows)
  - μια σχέση αποτελείται από διακεκριμένες πλειάδες
  - οι σχέσεις ορίζονται σαν σύνολα, επομένως δεν μπορούν να έχουν επαναλαμβανόμενα στοιχεία
  - η εφαρμογή του κανόνα αυτού συνεπάγεται κόστος για τον έλεγχο της μοναδικότητας των πλειάδων
  - πολλά συστήματα επιλέγουν τη μη-εφαρμογή του κανόνα αυτού για να αποφύγουν το επιπλέον κόστος

- Κλειδιά και υπερ-κλειδιά (keys and superkeys)
  - ο 3ος σχεσιακός κανόνας επιβάλει την απουσία πλειάδων οι οποίες έχουν τις ίδιες τιμές σε όλα τους τα γνωρίσματα
  - άρα πρέπει να υπάρχει ένα υποσύνολο του συνόλου των γνωρισμάτων το οποίο μπορεί να χρησιμοποιηθεί προκειμένου να διακρίνομε οποιεσδήποτε δύο πλειάδες μιας σχέσης
  - ένα σύνολο γνωρισμάτων μιας σχέσης αποτελεί ένα υπερ-κλειδί (superkey) της σχέσης αν οποιεσδήποτε δύο πλειάδες της σχέσης έχουν διακεκριμένες τιμές για το σύνολο αυτό των γνωρισμάτων
  - ένα σύνολο γνωρισμάτων μιας σχέσης αποτελεί κλειδί (key) της σχέσης αν το σύνολο αυτό είναι ένα υπερ-κλειδί και κανένα υποσύνολο αυτού δεν έχει αυτή την ιδιότητα

- Κλειδιά και υπερ-κλειδιά (keys and superkeys)
  - Η επιλογή των κλειδιών και υπερ-κλειδιών είναι θέμα της
     σχεδίασης μιας βάσης δεδομένων και δεν μπορεί να αποφασιστεί εξετάζοντας το περιεχόμενο της
  - Παράδειγμα

{pid, pname}: υπερ-κλειδί

{pid}: κλειδί

{pname}: δεν είναι κλειδί

#### **Products**

pid	pname	city	quantity	price
p01	comb	Dallas	111400	0.50
p02	brush	Newark	203000	0.50
p03	razor	Duluth	150500	1.00
p04	pen	Duluth	125000	1.00
p05	pencil	Dallas	221400	1.00

- Ορισμός Έστω R μια σχέση με σχήμα head(R) =  $\{A_1, A_2, ..., A_n\}$  και έστω t μια πλειάδα της R. Αν  $\{A_{i1}, A_{i2}, ..., A_{ik}\} \subseteq \{A_1, A_2, ..., A_n\}$ , τότε ο περιορισμός (restriction) t[ $A_{i1}, A_{i2}, ..., A_{ik}$ ] της t στα γνωρίσματα  $\{A_{i1}, A_{i2}, ..., A_{ik}\}$  είναι η πλειάδα τιμών της t στα γνωρίσματα  $A_{i1}, A_{i2}, ..., A_{ik}$
- Παράδειγμα: Αν t=(p01, comb, Dallas, 111400, 0.50) τότε
   t[pid,city,price]=(p01, Dallas, 0.50)
- Ορισμός Ενα κλειδί μιας σχέσης R με σχήμα  $head(R) = \{A_1, A_2, ..., A_n\}$  είναι ένα σύνολο γνωρισμάτων  $K = \{A_{i1}, A_{i2}, ..., A_{ik}\}$  για το οποίο:
  - $\blacksquare$  an u,v eίναι διακεκριμένες πλειάδες της R, τότε  $u[K] \neq v[K]$
  - δεν υπάρχει H ⊂ K με την παραπάνω ιδιότητα

- Ένα σύνολο γνωρισμάτων το οποίο πληροί την πρώτη ιδιότητα του προηγούμενου ορισμού λέγεται υπερ-κλειδί.
- Ένα κλειδί είναι ένα ελάχιστο υπερ-κλειδί.
- Το κενό σύνολο δεν μπορεί να είναι κλειδί.
- Μια σχέση μπορεί να έχει περισσότερα από ένα κλειδιά
  - τα κλειδιά μιας σχέσης ονομάζονται υποψήφια κλειδιά
  - ένα από τα υποψήφια κλειδιά επιλέγεται ως το πρωτεύον κλειδί
- Θεώρημα: Κάθε σχέση έχει τουλάχιστον ένα κλειδί
- Απόδειξη: Έστω ότι head(R) =  $\{A_1, A_2, ..., A_n\}$  και έστω  $S_1$  το σύνολο αυτό. Από τον σχεσιακό κανόνα 3, το  $S_1$  είναι υπερ-κλειδί. Άρα για κάποιες πλειάδες u,v με  $u \neq v$ , πρέπει να υπάρχει γνώρισμα  $A_i \in S_1$  τέτοιο ώστε  $u[A_i] \neq v[A_i]$

- Απόδειξη (συνέχεια):
  - Περίπτωση 1: S₁ είναι κλειδί
  - Περίπτωση 2: Υπάρχει S₂ ⊂ S₁ τέτοιο ώστε S₂ είναι υπερκλειδί. Αν το S₂ δεν είναι κλειδί, τότε πρέπει να υπάρχει S₃ ⊂ S₂ τέτοιο ώστε S₃ είναι υπερκλειδί. Συνεχίζοντας με αυτό τον τρόπο συμπεραίνομε ότι πρέπει να υπάρχει ένα σύνολο Sk το οποίο είναι υποσύνολο όλων των προηγούμενων συνόλων και το οποίο δεν έχει γνήσιο υποσύνολο το οποίο να είναι υπερκλειδί.

Aν  $\#S_i$  δηλώνει τον αριθμό των γνωρισμάτων στο σύνολο  $S_i$ , τότε  $\#S_1$ =n,  $\#S_i > \#S_{i+1}$  και  $\#S_k > 0$ .

Άρα, το  $S_k$  είναι κλειδί.

- Κενές Τιμές
- Παράδειγμα: Έστω το παρακάτω στιγμιότυπο της σχέσης Products. Θέλομε να εισάγομε ένα καινούργιο προϊόν αλλά δε γνωρίζομε όλη τη σχετική πληροφορία.

#### **Products**

pid	pname	city	quantity	price
p01	comb	Dallas	111400	0.50
p02	brush	Newark	203000	0.50
p03	razor	Duluth	150500	1.00
p04	pen	Duluth	125000	1.00
p05	pencil	Dallas	221400	1.00

- Κενές Τιμές
- Για τα γνωρίσματα των οποίων τις τιμές δεν γνωρίζομε, θα χρησιμοποιήσομε μια ειδική τιμή η οποία λέγεται κενή τιμή (null value)

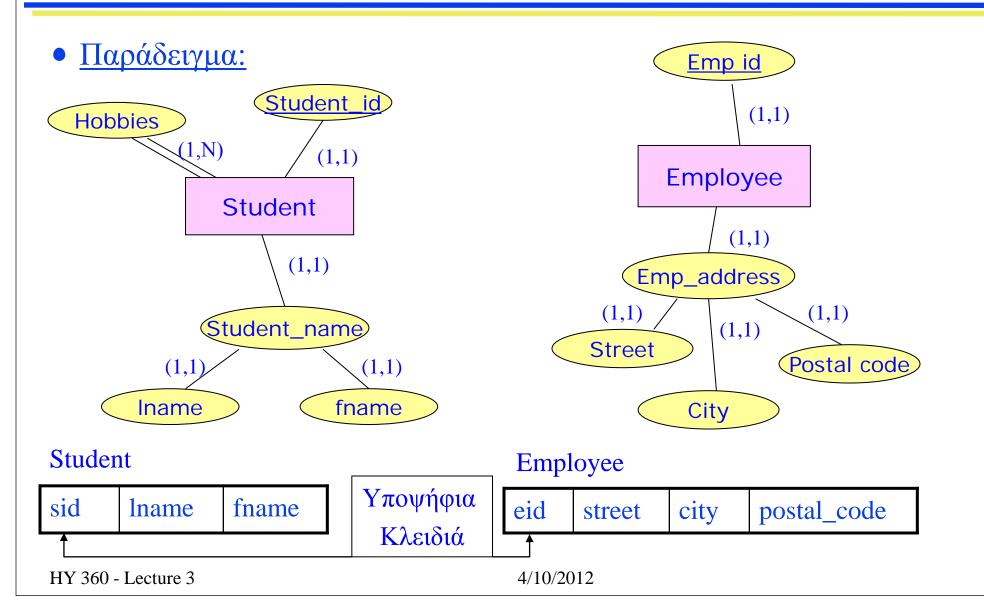
Η πλειάδα που εισάγεται έχει τη μορφή

pid	pname	city	quantity	price
p07	stapler	null	null	3.50

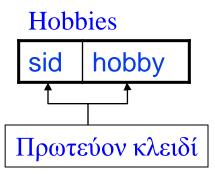
- Οι «άγνωστες» τιμές μπορεί να γίνουν γνωστές σε κάποια επόμενη κατάσταση.
- Σημασιολογία κενών τιμών:
  - υπαρκτή αλλά άγνωστη τιμή, ή
  - η εισαγωγή μιας τιμής δεν έχει νόημα για κάποια πλειάδα

- Κενές Τιμές
- Οι κενές τιμές γνωρισμάτων σχέσεων είναι διαφορετικές από τις αντίστοιχες «κενές τιμές» των τύπων δεδομένων (π.χ. null string)
- Προσοχή: ορισμένα γνωρίσματα δεν επιτρέπεται να δέχονται κενές τιμές
- Κανόνας 4: Ακεραιότητα Οντοτήτων (Entity Integrity)
  - γνωρίσματα τα οποία ανήκουν στο πρωτεύον κλειδί μιας σχέσης
     R δε δέχονται κενές τιμές για καμία πλειάδα της R
  - παραβίαση αυτού του κανόνα θα σήμαινε ότι δεν έχομε τρόπο να ανακτήσομε την πληροφορία που φυλάσσεται σε πλειάδες με κενές τιμές

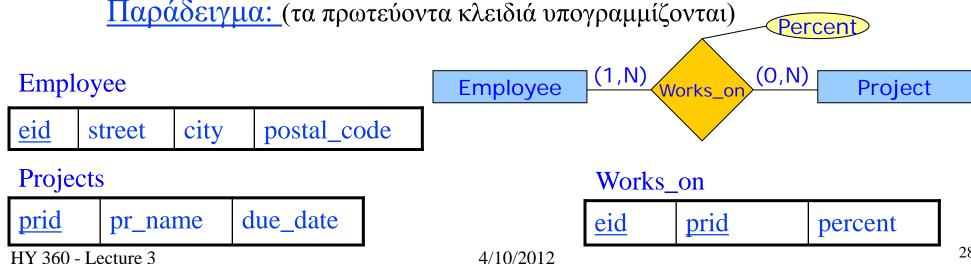
- Ένα εννοιολογικό μοντέλο που εκφράζεται με το συμβολισμό του μοντέλου Οντοτήτων Σχέσεων μπορεί να μετατραπεί σε ένα σύνολο σχέσεων σύμφωνα με τους παρακάτω κανόνες:
  - 1. Κάθε οντότητα σε ένα διάγραμμα Ε-R απεικονίζεται σε μια σχέση με το ίδιο όνομα.
    - Ολα τα μονότιμα γνωρίσματα της οντότητας (απλά και σύνθετα) γίνονται γνωρίσματα της αντίστοιχης σχέσης.
    - Τα αναγνωριστικά γνωρίσματα της σχέσης σχηματίζουν ένα υποψήφιο κλειδί της σχέσης
    - Τα στιγμιότυπα της οντότητας γίνονται πλειάδες της σχέσης



- 2. Μια οντότητα Ε με αναγνωριστικό p και ένα γνώρισμα πολλαπλών τιμών a απεικονίζεται σε μια σχέση με όνομα το όνομα του γνωρίσματος πολλαπλών τιμών.
  - Το σχήμα της σχέσης περιλαμβάνει τα γνωρίσματα p και a και οι πλειάδες της είναι ζεύγη τιμών των p και a
  - Το πρωτεύον κλειδί της σχέσης είναι το σύνολο που περιλαμβάνει τα γνωρίσματα p και a.
  - Παράδειγμα:



Μια Ν-Ν σχέση r μεταξύ οντοτήτων Ε και F απεικονίζεται σε μια σχέση R το σχήμα της οποίας περιέχει όλα τα γνωρίσματα που ανήκουν στα πρωτεύοντα κλειδιά των σχέσεων που αντιστοιχούν στις οντότητες Ε και F. Ο συνδυασμός αυτός σχηματίζει το πρωτεύον κλειδί της R. Επίσης, το σχήμα της R περιέχει όλα τα γνωρίσματα της σχέσης r. Οι πλειάδες της R αντιστοιχούν στα στιγμιότυπα της r.



4. Για μια N-1σχέση r μεταξύ οντοτήτων E και F δε δημιουργούμε νέα σχέση για την αναπαράστασή της. Αν max-card(F,R) = 1, τότε η σχέση της F πρέπει να περιέχει γνωρίσματα που αντιστοιχούν στο πρωτεύον κλειδί της E (ξένο κλειδί). Αν η F έχει υποχρεωτική συμμετοχή στην r, τότε το ξένο κλειδί δε δέχεται κενές τιμές.

Παράδειγμα:

Instructor



Course

Instructor

iid lname fname

Course

cid cname credits iid

Ξένο κλειδί

29

**Project** 

## Μοντέλο Οντοτήτων – Σχέσεων $\rightarrow$ Σχεσιακό Μοντέλο

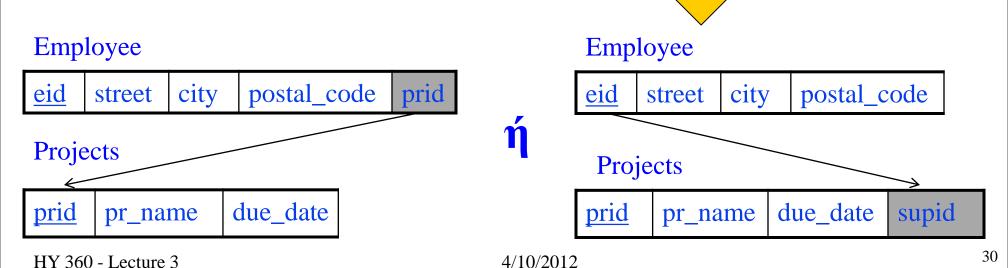
5. Αν οι οντότητες Ε και F συμμετέχουν σε μια 1-1 σχέση r, τότε αν η συμμετοχή των οντοτήτων είναι προαιρετική, δημιουργούμε σχέσεις για τις Ε και F και προσθέτομε στη μία από αυτές ένα γνώρισμα για το πρωτεύον κλειδί της άλλης. Αν η συμμετοχή τους είναι υποχρεωτική, τότε οι δύο σχέσεις μπορούν να συνδυαστούν σε μία.

(0,1)

Supervises

(0,1)

Παράδειγμα:



**Employee** 

- 6. Στην περίπτωση σχέσεων εξειδίκευσης (IsA), υπάρχουν δύο περιπτώσεις:
  - Αν κάθε στιγμιότυπο της πατρικής οντότητας είναι υποχρεωτικά και στιγμιότυπο μιας μόνο από τις εξειδικευμένες οντότητες, τότε δημιουργούμε μία σχέση μόνο για κάθε εξειδικευμένη οντότητα και όχι για την πατρική. Οι σχέσεις αυτές περιέχουν όλα τα γνωρίσματα της πατρικής.

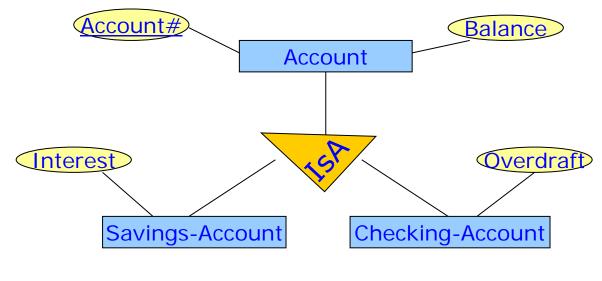
#### Παράδειγμα:

Savings-Account

acc\_no balance interest

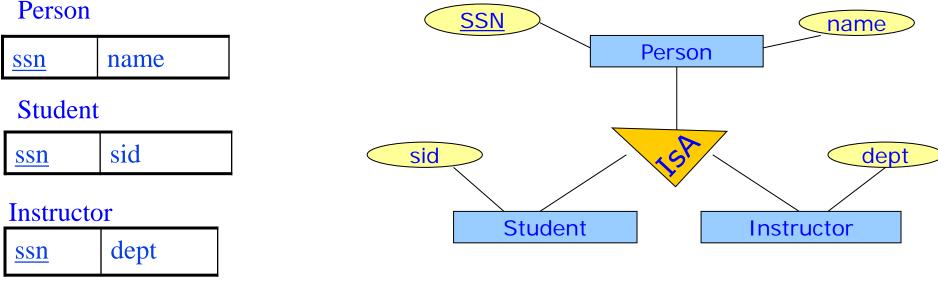
Checking-Account

acc\_no balance overdraft



• Στην αντίθετη περίπτωση, όταν δηλαδή υπάρχουν στιγμιότυπα της πατρικής οντότητας που μπορεί να μην αντιστοιχούν σε στιγμιότυπο κάποιας από τις εξειδικευμένες οντότητες (ή αντιστοιχούν σε περισσότερες από μία) δημιουργούμε μια σχέση για την πατρική οντότητα, και μια για κάθε εξειδικευμένη, προσθέτοντας σ' αυτές το πρωτεύον κλειδί της πατρικής.

#### Παράδειγμα:



32