

Αποσύνθεση χωρίς Απώλεια Πληροφορίας

(Lossless-Join Decomposition)

- Η κανονικοποίηση σχημάτων σχεσιακών ΒΔ εξαρτάται από τη δυνατότητα αποσύνθεσης σχημάτων σε "μικρότερα" αποφεύγοντας συγχρόνως ανωμαλίες ενημέρωσης.
- Δεδομένης μιας σχέσης R , μια **αποσύνθεση** (*decomposition*) της R σε k σχέσεις είναι ένα σύνολο $\{R_1, R_2, \dots, R_k\}$ τέτοιο ώστε $Head(R) = \bigcup_{i=1}^k Head(R_i)$. Δεδομένου ενός στιγμιότυπου της R , το περιεχόμενο κάθε μιας από τις σχέσεις R_i καθορίζεται από την προβολή των πλειάδων της R στα γνωρίσματα της R_i .

- Μια αποσύνθεση $\{R_1, R_2, \dots, R_k\}$ της σχέσης R με συναρτησιακές εξαρτήσεις F λέγεται αποσύνθεση χωρίς απώλεια πληροφορίας (lossless-join decomposition) αν, ανεξάρτητα από το περιεχόμενο της R , οι συναρτησιακές εξαρτήσεις εξασφαλίζουν ότι $R = R_1 \bowtie R_2 \bowtie \dots \bowtie R_k$.
- Σε μια αποσύνθεση χωρίς απώλεια πληροφορίας, μπορούμε πάντα να ανακατασκευάσουμε την αρχική σχέση από τον υπολογισμό της συνένωσης (Join) των σχέσεων που προκύπτουν από την αποσύνθεση.
- Διαφορετικά, η συνένωση των σχέσεων μπορεί να δώσει πληροφορία η οποία δεν υπήρχε στην αρχική σχέση.

Παράδειγμα: Αποσύνθεση με απώλεια πληροφορίας

R (ABC)

A	B	C
a_1	100	c_1
a_2	200	c_2
a_3	300	c_3
a_4	200	c_4

R1 (AB)

A	B
a_1	100
a_2	200
a_3	300
a_4	200

R2 (BC)

B	C
100	c_1
200	c_2
300	c_3
200	c_4

R1 ⋈ R2

A	B	C
a_1	100	c_1
a_2	200	c_2
a_2	200	c_4
a_3	300	c_3
a_4	200	c_2
a_4	200	c_4

Εξετάζοντας μόνο τις $R1$ και $R2$ δεν μπορούμε να πούμε από ποιά σχέση προήλθαν. Θα μπορούσαν να έχουν προέλθει και από την

A	B	C
a_1	100	c_1
a_2	200	c_2
a_2	200	c_4
a_3	300	c_3
a_4	200	c_4

Η απώλεια πληροφορίας προήλθε από τις πλειάδες $(a_2, 200, c_4)$ και $(a_4, 200, c_2)$. Εμφανίζονται στην $R1 \bowtie R2$ επειδή έχουν κοινή τιμή στο γνώρισμα B αλλά όχι στην αρχική σχέση.

Παράδειγμα: Αποσύνθεση με απώλεια πληροφορίας

R (ABC)

A	B	C
a_1	100	c_1
a_2	200	c_2
a_3	300	c_3

R1 (AB)

A	B
a_1	100
a_2	200
a_3	300

R2 (BC)

B	C
100	c_1
200	c_2
300	c_3

$$R1 \bowtie R2 = R$$

Τι θα συμβεί όμως αν το περιεχόμενο της R αλλάξει και προστεθεί μια πλειάδα με τιμή που ήδη υπάρχει στο κοινό γνώρισμα?

- Δεν μπορούμε να κρίνουμε αν μια αποσύνθεση πάσχει από απώλεια πληροφορίας εξετάζοντας μόνο το περιεχόμενο των σχέσεων.
- Χρειάζονται επιπλέον κανόνες οι οποίοι λαμβάνουν υπόψη τους τις συναρτησιακές εξαρτήσεις.

Παράδειγμα: Εστω ότι η εξάρτηση $B \rightarrow C$ ισχύει για την σχέση R .

R (ABC)

A	B	C
a_1	100	c_1
a_2	200	c_2
a_3	300	c_3

R1 (AB)

A	B
a_1	100
a_2	200
a_3	300

R2 (BC)

B	C
100	c_1
200	c_2
300	c_3

Η εισαγωγή της πλειάδας $(a_4, 200, c_4)$ αποτυγχάνει γιατί ισχύει η εξάρτηση $B \rightarrow C$. Η εισαγωγή της πλειάδας $(a_4, 200, c_2)$ επιτρέπεται.

Η ακόλουθη αποσύνθεση δεν πάσχει από απώλεια πληροφορίας.

R (ABC)

A	B	C
a_1	100	c_1
a_2	200	c_2
a_3	300	c_3
a_4	200	c_2

R1 (AB)

A	B
a_1	100
a_2	200
a_3	300
a_4	200

R2 (BC)

B	C
100	c_1
200	c_2
300	c_3

Το ακόλουθο θεώρημα προσδιορίζει τη σχέση μεταξύ κλειδιών σχέσεων και συναρτησιακών εξαρτήσεων.

Θεώρημα: Εστω μια σχέση R και ένα σύνολο γνωρισμάτων $X \subseteq \text{Head}(R)$. Τότε οι ακόλουθες προτάσεις είναι ισοδύναμες: (1) X είναι κλειδί της R ; (2) X προσδιορίζει συναρτησιακά όλα τα γνωρίσματα στην R .

Απόδειξη:

(1) \Rightarrow (2): αν το X είναι κλειδί, τότε δεν μπορούν να υπάρχουν στην R πλειάδες οι οποίες συμφωνούν σε όλα τα γνωρίσματα του X . Αρα, δεν είναι δυνατόν να υπάρχουν πλειάδες οι οποίες συμφωνούν στις τιμές των γνωρισμάτων X και δεν συμφωνούν στα υπόλοιπα γνωρίσματα.

Εξ'ορισμού η συναρτησιακή εξάρτηση $X \rightarrow \text{Head}(R)$ ισχύει.

(2) \Rightarrow (1): αν ισχύει η $X \rightarrow Head(R)$, τότε δεν είναι δυνατόν να υπάρχουν δύο πλειάδες οι οποίες συμφωνούν στα γνωρίσματα X και δεν συμφωνούν στα υπόλοιπα γνωρίσματα. Αρα, είτε δύο πλειάδες οι οποίες συμφωνούν στα γνωρίσματα X θα συμφωνούν και στα υπόλοιπα γνωρίσματα, είτε όλες οι πλειάδες έχουν διακεκριμένες τιμές στα γνωρίσματα του X . Καθώς δεν επιτρέπεται να υπάρχουν επαναλαμβανόμενες πλειάδες σε μια σχέση, και οι δύο αυτές περιπτώσεις μας οδηγούν στο συμπέρασμα ότι το X είναι κλειδί της R .

Αρα, (1) \equiv (2).

Μια ικανή συνθήκη για έλεγχο της ιδιότητας της μη-απώλειας πληροφορίας για μια αποσύνθεση:

Θεώρημα: Δεδομένης μιας σχέσης R και ενός συνόλου ΣΕ F οι οποίες πληρούνται στην R , μια αποσύνθεση της R στις σχέσεις R_1 και R_2 δεν πάσχει από απώλεια πληροφορίας αν τουλάχιστον μία από τις ακόλουθες ΣΕ είναι λογική συνέπεια των ΣΕ στο F :

$$(1) \text{Head}(R_1) \cap \text{Head}(R_2) \rightarrow \text{Head}(R_1)$$

$$(2) \text{Head}(R_1) \cap \text{Head}(R_2) \rightarrow \text{Head}(R_2)$$

Παράδειγμα: Εστω ότι η ΣΕ $B \rightarrow C$ ισχύει στη σχέση $R(ABC)$. Η R αποσυντίθεται στις $R_1(AB)$ και $R_2(BC)$. Εξετάστε αν η αποσύνθεση πάσχει από απώλεια πληροφορίας.

$$Head(R_1) \cap Head(R_2) = B$$

Πρέπει να δείξουμε ότι ισχύει μια από τις ΣΕ (1) $B \rightarrow AB$ και (2) $B \rightarrow BC$.

Από την $B \rightarrow C$ εξάγεται η $B \rightarrow BC$ με χρήση του κανόνα επαύξησης.

Αρα, η αποσύνθεση δεν πάσχει από απώλεια πληροφορίας.

Πόρισμα: Αν $\{R_1, R_2\}$ είναι μια αποσύνθεση χωρίς απώλεια πληροφορίας για τη σχέση R και $\{R_3, R_4\}$ είναι μια αποσύνθεση χωρίς απώλεια πληροφορίας της σχέσης R_2 , τότε $\{R_1, R_3, R_4\}$ είναι μια αποσύνθεση χωρίς απώλεια πληροφορίας της R .

Παράδειγμα: Αποσύνθεση της σχέσης emp_info.

emp_info

emp_id	emp_name	...	skill_id	skill_name	skill_date	skill_lvl
--------	----------	-----	----------	------------	------------	-----------

Συναρτησιακές Εξαρτήσεις :

1. emp_id \rightarrow emp_name emp_phone dept_name
2. dept_name \rightarrow dept_phone dept_mgrname
3. skill_id \rightarrow skill_name
4. emp_id skill_id \rightarrow skill_date skill_lvl

Η ακόλουθη αποσύνθεση δεν πάσχει από απώλεια πληροφορίας:

emps

emp_id	emp_name	emp_phone	dept_name
--------	----------	-----------	-----------

depts

dept_name	dept_phone	dept_mgrname
-----------	------------	--------------

emp-skills

emp_id	skill_id	skill_date	skill_lvl
--------	----------	------------	-----------

skills

skill_id	skill_name
----------	------------

Κανονικές Μορφές (Normal Forms)

- Παρέχουν ένα τυπικό πλαίσιο για ανάλυση σχεσιακών σχημάτων βασισμένη στον ορισμό κλειδιών και συναρτησιακών εξαρτήσεων.
- Σχεσιακά σχήματα που ανήκουν σε συγκεκριμένες κανονικές μορφές έχουν ορισμένες επιθυμητές ιδιότητες (π.χ. έλλειψη ανωμαλιών ενημέρωσης).
- Οι κανονικές μορφές καθορίζονται με βάση κάποιες συνθήκες οι οποίες αν δεν πληρούνται έχουν ως αποτέλεσμα την αποσύνθεση σχημάτων σε σχέσεις οι οποίες πληρούν τις κανονικές μορφές.

Ορισμός: Δεδομένης μιας καθολικής σχέσης R , ενός συνόλου F από ΣΕ και μιας αποσύνθεσης $\{R_1, \dots, R_k\}$ της R , η ΣΕ $X \rightarrow Y$ του F **διατηρείται** σε κάποια σχέση R_i της αποσύνθεσης αν και μόνο αν $X \cup Y \subseteq \text{Head}(R_i)$.

- Η διατήρηση των ΣΕ είναι μια επιθυμητή ιδιότητα:
αν μια ΣΕ $X \rightarrow Y$ διατηρείται σε μια σχέση R_i , τότε η επαλήθευση της ΣΕ μετά από κάποια ενημέρωση της σχέσης μπορεί να γίνει τοπικά, δηλαδή χωρίς να χρειάζεται να υπολογιστεί κάποιο \bowtie .
- Μας ενδιαφέρουν αποσυνθέσεις σε κανονικές μορφές χωρίς απώλεια πληροφορίας και με διατήρηση των ΣΕ.

Παραάδειγμα: Κάθε μια από τις ΣΕ

1. $\text{emp_id} \rightarrow \text{emp_name emp_phone dept_name}$
2. $\text{dept_name} \rightarrow \text{dept_phone dept_mgrname}$
3. $\text{skill_id} \rightarrow \text{skill_name}$
4. $\text{emp_id skill_id} \rightarrow \text{skill_date skill_lvl}$

διατηρείται σε κάποια από τις σχέσεις του σχήματος:

emps

emp_id	emp_name	emp_phone	dept_name
--------	----------	-----------	-----------

skills

skill_id	skill_name
----------	------------

Κανονική Μορφή Boyce-Codd

Μια σχέση R σε ένα σχεσιακό σχήμα με ΣΕ F είναι σε κανονική μορφή Boyce-Codd (BCNF) αν ισχύει η παρακάτω ιδιότητα:

για οποιαδήποτε ΣΕ $X \rightarrow A$ του F^+ η οποία διατηρείται στην R και για την οποία $A \notin X$, το X είναι κλειδί της R .

- Ένα σχεσιακό σχήμα είναι σε κανονική μορφή Boyce-Codd αν κάθε σχέση του είναι σε κανονική μορφή Boyce-Codd.
- Αν μια σχέση είναι σε BCNF αυτό σημαίνει ότι κανένα γνώρισμα της σχέσης δεν εξαρτάται συναρτησιακά από κανένα υποσύνολο των γνωρισμάτων της σχέσης το οποίο δεν είναι κλειδί.

Παράδειγμα: Η συναρτησιακή εξάρτηση

$\text{emp_id skill_id} \rightarrow \text{skill_date skill_lvl}$

διατηρείται στη σχέση

emp-skills

emp_id	skill_id	skill_date	skill_lvl
--------	----------	------------	-----------

Από την $\text{emp_id skill_id} \rightarrow \text{skill_date skill_lvl}$ εξάγονται με χρήση του κανόνα της αποσύνθεσης οι ΣΕ

$\text{emp_id skill_id} \rightarrow \text{skill_date}$

$\text{emp_id skill_id} \rightarrow \text{skill_lvl}$

Και οι δύο διατηρούνται στη σχέση **emp-skills**. Δεν υπάρχει άλλη εξάρτηση η οποία διατηρείται στη σχέση αυτή. Το μοναδικό κλειδί της σχέσης είναι το ζεύγος (**emp_id, skill_id**). Αρα, η σχέση είναι σε BCNF.

Παρόμοια, η εξάρτηση $\text{emp_id} \rightarrow \text{emp_name emp_phone dept_name}$ διατηρείται στη σχέση

emps

emp_id	emp_name	emp_phone	dept_name
--------	----------	-----------	-----------

Από αυτή την εξάρτηση, εξάγονται οι ακόλουθες ΣΕ

$\text{emp_id} \rightarrow \text{emp_name}$

$\text{emp_id} \rightarrow \text{emp_phone}$

$\text{emp_id} \rightarrow \text{dept_name}$

οι οποίες όλες διατηρούνται στη σχέση **emps**. Καμία άλλη ΣΕ δεν διατηρείται στη σχέση αυτή.

Το μοναδικό κλειδί της σχέσης είναι το γνώρισμα **emp_id**. Άρα η σχέση είναι σε BCNF.

Με παρόμοιο τρόπο μπορούμε να δείξουμε ότι ολόκληρο το σχήμα είναι σε BCNF.

Εστω τώρα ότι θέλουμε να προσθέσουμε την εξάρτηση
 $\text{dept_mgrname} \rightarrow \text{dept_name}$. Συνεχίζει το σχήμα να είναι σε BCNF?

Οι ΣΕ οι οποίες διατηρούνται είναι:

$\text{dept_name} \rightarrow \text{dept_phone}$

$\text{dept_mgrname} \rightarrow \text{dept_name}$

$\text{dept_mgrname} \rightarrow \text{dept_phone}$

Αρα, το σχήμα συνεχίζει να είναι σε BCNF.

- Η κανονική μορφή BCNF είναι πολύ περιοριστική: δεν είναι πάντα δυνατόν να βρεθεί μια αποσύνθεση σε BCNF η οποία ταυτόχρονα να μην πάσχει από απώλεια πληροφορίας και να διατηρεί τις συναρτησιακές εξαρτήσεις.
- Λιγότερο περιοριστικές κανονικές μορφές μας επιτρέπουν να ορίζουμε αποσυνθέσεις σχημάτων χωρίς απώλεια πληροφορίας και με διατήρηση των εξαρτήσεων.

