
ΕΙΣΑΓΩΓΗ

Π. ΒΑΣΙΛΕΙΑΔΗΣ

Η ιστορία των βάσεων δεδομένων ξεκινάει περίπου 30 χρόνια πριν από σήμερα. Για την ακρίβεια, τα πρώτα δείγματα βάσεων δεδομένων, ή αλλιώς "τραπεζών δεδομένων" (data banks), όπως ονομάζονταν τότε, μπορούν να βρεθούν στα τέλη της δεκαετίας του '60. Εκείνη την εποχή τυποποιήθηκαν οι δικτυακές βάσεις δεδομένων (network databases), οι οποίες είναι γνωστές και με το όνομα βάσεις δεδομένων CODASYL, από το όνομα της επιτροπής που τις τυποποίησε. Επίσης, την εποχή εκείνη πρωτοεμφανίζεται και το IMS, το πιο γνωστό σύστημα διαχείρισης ιεραρχικών βάσεων δεδομένων (hierarchical DBMS). Το IMS, που είναι προϊόν της IBM, είναι ακόμα και σήμερα, ένα από τα συστήματα βάσεων δεδομένων με τις περισσότερες εγκαταστάσεις. Τα δύο μοντέλα αυτά χαρακτηρίζονται από το γεγονός ότι (α) η διαχείριση των δεδομένων γίνεται σε χαμηλό επίπεδο (με αποτέλεσμα όλες οι εφαρμογές να εξαρτώνται άμεσα από την τοποθεσία αποθήκευσης των δεδομένων) και (β) διαχειριζόμαστε τα δεδομένα ατομικά (record-at-a-time).

Στα 1970, όμως, κάτι αλλάζει: ο E. F. Codd δημοσιεύει το άρθρο "*A Relational Model of Data for Large Shared Data Banks*", στο περιοδικό "Communications of the ACM". Το εν λόγω άρθρο, θεωρείται ως η εργασία με τη μεγαλύτερη επίδραση στο χώρο των βάσεων δεδομένων, καθώς είναι αυτή που εισήγαγε το σχεσιακό μοντέλο βάσεων δεδομένων.

Το σχεσιακό μοντέλο ήταν μια επαναστατική νέα στροφή στο νέο -την εποχή εκείνη- χώρο των βάσεων δεδομένων. Το μοντέλο αυτό -σε αντιδιαστολή με τα προηγούμενα- χαρακτηρίζεται από τη δυνατότητα μαζικής (set-at-a-time) επεξεργασίας δεδομένων και από την ανεξαρτησία των εφαρμογών από τη φυσική υλοποίηση. Το σχεσιακό μοντέλο, αφενός χαρακτηρίζεται από ένα πολύ ισχυρό θεωρητικό υπόβαθρο και αφετέρου από τη δυνατότητα άμεσης υλοποίησης: είναι μια από τις λίγες θεωρητικές εργασίες με τόσο μεγάλη πρακτική επίδραση. Ο E. F. Codd, μια δεκαετία αργότερα, το 1981, έλαβε το Turing Award για την συμβολή του αυτή.

Φυσικά, μέχρι να πάρει το Turing Award ο Codd, έπρεπε το μοντέλο του να περάσει το τεστ του χρόνου. Ο πόλεμος ξεκίνησε αμέσως, καθώς οι προηγούμενες τεχνολογίες είχαν τα

επιχειρήματά τους: παρείχαν ώριμα συστήματα, με εξαιρετική απόδοση σε πολλές περιπτώσεις και οι προγραμματιστές ήταν εξοικειωμένοι με αυτά (χαρακτηριστικά που έχουν διατηρήσει μέχρι και σήμερα, και τα οποία τους επιτρέπουν να επιβιώνουν ανετότατα, 30 χρόνια αργότερα από την εποχή που πρωτοεμφανίστηκαν). Γύρω στα μέσα τις δεκαετίας του '70, όμως, η μάχη κρίθηκε με την ανάπτυξη δύο σχεσιακών συστημάτων (που είναι οι πρόγονοι όλων των εμπορικών συστημάτων σήμερα): της *Ingres* από το Πανεπιστήμιο της Καλιφόρνιας και του *System R* της IBM από το ερευνητικό κέντρο του San Jose (το οποίο τώρα ονομάζεται IBM Almaden Research Center).

Η *Ingres*, κατασκευάστηκε από μια ομάδα ερευνητών με επικεφαλής τον M. Stonebraker. Σήμερα υπάρχουν διάφοροι απόγονοι της αρχικής αυτής προσπάθειας: *Ingres*, *Sybase* (και ως εκ τούτου και ο *Microsoft SQL Server*), *Britton-Lee/Sharebase*, *PACE*. Ακόμα, σημαντική επιρροή είχε και στην αρχιτεκτονική της *Informix*.

Το *System R*, από την άλλη πλευρά, φτιάχτηκε από ερευνητές της IBM (Irv Traiger, Bruce Lindsay, Paul McJones, Mike Blasgen, Mario Schkolnick, Bob Selinger, Bob Yost, Jim Gray), ήταν επίσης πρόγονος για διάφορα σχεσιακά συστήματα που υπάρχουν σήμερα: *Oracle*, *HP Allbase*, *IBM SQL/DS*, *IBM DB2*, *Tandem Non-Stop SQL*.

Και τα δύο συστήματα, όντας ανταγωνιστικά μεταξύ τους, κυριάρχησαν στον εμπορικό, αλλά και τον ερευνητικό χώρο των σχεσιακών συστημάτων, μέχρι τις αρχές της δεκαετίας του '80, όταν η *Oracle* παίρνει τα πρωτεία από την *IBM*. Στα μέσα της δεκαετίας του '80, έχουμε το πρώτο στάνταρτ για μια γλώσσα διαχείρισης και επερώτησης σχεσιακών βάσεων δεδομένων, την *Structured Query Language - SQL*!

Η συνέχεια, τουλάχιστον στον εμπορικό χώρο, είναι γνωστή: τα σχεσιακά συστήματα είναι τα μόνα που πουλάνε πλέον, οι προγραμματιστές είναι, εν γένει, πιο ευτυχείς, πλην όμως τα προβλήματα (και ως εκ τούτου και οι τεχνολογικές προκλήσεις) δεν έχουν σταματήσει να υφίστανται...

Το σχεσιακό μοντέλο ήταν ένα πλήρες και καλά ορισμένο μοντέλο. Οι υλοποιήσεις του, όμως, έγιναν με κάποιους "συμβιβασμούς". Το αποτέλεσμα ήταν μάλλον μακροπρόθεσμο, σε σχέση με την εκφραστικότητα και την ορθή σημασιολογία της γλώσσας ερωτήσεων. Διάφορες προσπάθειες έγιναν και συνεχίζουν να γίνονται, ώστε να κάνουν την *SQL* (και τα σχεσιακά συστήματα) πιο ισχυρή υπολογιστικά, φιλική και ορθά συμπεριφερόμενη. Το δεύτερο στάνταρτ της *SQL*, βγήκε στα 1992 (*SQL-92*) και το τρίτο (ονόματι *SQL-3*) είναι ήδη σε διαδικασία παραγωγής.

Παράλληλα, νέες τεχνολογίες ήρθαν να καλύψουν το κενό: η πιο σημαντική από τις προσπάθειες που έγιναν, ήταν η εισαγωγή των αντικειμενοστρεφών συστημάτων βάσεων δεδομένων, γύρω στα μέσα της δεκαετίας του '80. Τα αντικειμενοστρεφή συστήματα στηρίζονταν στο ομώνυμο μοντέλο δεδομένων που διέφερε από το σχεσιακό και επανέφερε στο προσκήνιο κάποια από τα πλεονεκτήματα των δικτυακών βάσεων δεδομένων. Είναι, δε, χαρακτηριστικό, ότι το τελευταίο στάνταρτ της *SQL*, είναι αφιερωμένο, στο μεγαλύτερο μέρος του, σε αντικειμενοστρεφείς επεκτάσεις στη σχεσιακή τεχνολογία.

Οι τεχνολογικές προκλήσεις για έρευνα (αλλά και πρακτική εφαρμογή) ενισχύονται από την τάση ενσωμάτωσης στην τεχνολογία των βάσεων δεδομένων, όλων των χαρακτηριστικών που συναντώνται σε διάφορες εφαρμογές. Με βάση τα δύο μοντέλα που προαναφέραμε, η έρευνα στο χώρο των βάσεων δεδομένων είχε διάφορα ενδιαφέροντα αποτελέσματα σε σημαντικούς τομείς εφαρμογών:

- τα γεωγραφικά συστήματα πληροφοριών συνεργάζονται με *βάσεις χωρικών δεδομένων*,

- οι εφαρμογές πραγματικού χρόνου με *ενεργές και χρονικές βάσεις δεδομένων*,
- οι εφαρμογές διαχείρισης γνώσης με *επαγωγικές βάσεις δεδομένων*,
- τα συστήματα λήψεως αποφάσεων με εφαρμογές *αποθηκών δεδομένων*

... και όλα αυτά σε περιβάλλοντα όπου η πληροφορία μπορεί να είναι είτε συγκεντρωτικά αποθηκευμένη είτε γεωγραφικά κατανεμημένη και επεξεργαζόμενη με συμβατικές ή με παράλληλες αρχιτεκτονικές. Για όλα αυτά τα ζητήματα, θα μιλήσουμε στα επόμενα κεφάλαια αυτού του βιβλίου...

Για την ιστορία που μόλις διηγήθήκαμε, στηριχθήκαμε κυρίως στο [Hell98].

- [Codd70] E. F. Codd, "A Relational Model of Data for Large Shared Data Banks", Communications of the ACM, 13(6), 1970.
- [Hell98] J. Hellerstein, Lecture Notes on Implementation of Database Systems, Univ. of California, College of Engineering, Department of EECS, Computer Science Division, CS286, Spring 1996. Available at:
<http://epoch.cs.berkeley.edu:8000/personal/jmh/cs286/lec1.html>

