

## אוגדן בעיות מפורסמות במדעי המחשב

### מסדי נתונים טבלאיים ויישומים עיסקיים

#### כתבה שרה פולק

#### תיאור הבעיה:

שימוש במחשב בארגונים עסקיים הוא בעיקר לשמירה ואחזור של מידע מבסיס נתונים המורכב מכמות גדולה של נתונים מסוגים שונים. לדוגמה: מערכת לניהול קופות בסופרמרקט צריכה לעקוב אחר כל מוצר, מערכת הזמנת טיסות צריכה להשתמש בהרבה מקומות בו זמנית כדי למקם נוסעים על טיסות רבות בתאריכים שונים. מערכת ספריה מאחסנת אלפי ספרים ומאפשרת איתור מהיר של פרסום בחתכים שונים. מנועי חיפוש באינטרנט סורקים מיליונים של דפי מידע כדי להוציא תגובות מיידיות לשאלות מגוונות. למעשה מיליוני עסקים קטנים וגדולים, פרטיים וארגונים ממשלתיים, משתמשים בבסיסי נתונים כדי לעקוב אחר כל דבר החל ממלאי ועד לרצף DNA של אדם. בעיה נוספת הקשורה לתחום זה נובעת מכך שארגונים עסקיים הם דינמיים והצרכים שלהם משתנים במשך הזמן, ובהתאם הדרישות מאותם נתונים משתנות גם הן, כך שבסיס הנתונים צריך לאפשר שימוש גמיש בהתאם. בבעיות מסוג זה מטפלים כיום במערכות שמבוססות על בסיס נתונים יחסי שהומצא בשנות ה-1960 – 1970. אבל זה לא התחיל כך.

#### רקע

ממשלת ארה"ב היא מאז ותמיד הזדקקה לאמצעים שיסייעו להפיק אוסף עצום של דוחות מסודר וממוין בהתאם לצרכיה. ב-1980 המציא Hollerith הציוד הראשון שהומצא לעיבוד מידע אוטומטי והיא מכונת לניקוב הכרטיסים. השימוש במכונה היה רב וב-1911 התמזגה החברה של הולרייט עם חברה נוספת וקמה חברת IBM שראשי התיבות שלה מצביעות על הייעוד שהיה לה - International Business Machines. במהלך מלחמת העולם הראשונה ממשלת ארה"ב השתמשה בציוד המכני לעיבוד של נתונים רבים שהיו דרושים לניהול ייצור תעשייתי ולאיסוף מסות של נתונים כדי לעבדם לדוחות. הדרישה לעיבודים אוטומטיים רק הלכה וגברה, וכמענה פותח ב-IBM המחשב הגדול הראשון UNIVAC 1, בו השתמשו משרדי ממשלה רבים בארה"ב. באותם הזמנים, מידע אוסון במבנה של קבצים שהיה תלוי באמצעי האחסון הפיזי (טייפים, דיסקים וכדומה). רק בשנות 1960 המוקדמות, הועלה הרעיון שמידע יכול להיות מאוחסן מובנה ומתופעל בנפרד ובלי קשר לאמצעי הפיזי עליו הוא מאוחסן ובמכונה בה הוא מעובד. באותה תקופה נקבעה שפת COBOL כשפה לעיבוד קבצי נתונים שאוחסנו אז על אמצעי אחסון שהיוו את המילה האחרונה – דיסק מגנטי ולא טייפ מגנטי. אמצעי אחסון אלו שאפשרו גישה ישירה לנתונים ולא רק גישה סדרתית. בהתאם החברה לתקינה CodeasyI קבעה סטנדרטים לארגון וניהול נתונים על בסיס מודל שנקרא מודל רשתי. במקביל IBM, החברה המתחרה, הוציאה מערכת לניהול בסיס נתונים במודל אחר מודל היררכי.

בשנים אלו עבד Codd ב-IBM, והוא חשב שהמודלים הקיימים לניהול בסיסי נתונים, קרי המודל הרשתי והמודל ההיררכי, הם בעייתיים ויש לשפרם. מודל היררכי היה מבוסס על מבנה נתונים עץ וקשה לייצג קשרים מסוג רבים לרבים, ומודל רשתי אפשר קישור רבים לרבים בין נתונים אך דרש זמן עיבוד רב.

במאמר פורץ דרך בתחום בשם "A Relational Model of Data for Large Shared Data Banks" שכתב Codd ב-1970, הוא תיאר דרך חדשה לארגון בסיס נתונים וניהולם הוא כינה אותו בשם מודל יחסי (Relational model). מודל זה מבוסס על שני עקרונות:

- אחסון נתונים בצורה בלתי תלויה באמצעי האחסון.
- גישה לנתונים בשפה לא פרוצדורלית.

לפי תפישתו של Codd, משתמש ביישומי מחשב לא צריך לדעת את המבנה של הנתונים כדי לשאול שאילתא ולאחזר מידע. Codd הציע מודל המבוסס על 12 עקרונות לבניית מסד נתונים יחסי.

ההצעה של Codd לא זכתה בתחילה לגיבוי, בלשון המעטה ב-IBM. האינטרסים הם כמו תמיד כסף: IBM שיווק את מע' לניהול בסיסי נתונים IMS וטענה שהיא מוצלחת בעוד שרעיונו של קוד עוד לא מבוסס (מוכר מה שעובד לא נוגעים). אבל Codd התעקש והמשיך לפרסם את רעיונותיו, רעיונות שהחלו לעניין הן את המדענים והן את ממשלת ארה"ב. בשנות ה-1970 החלו במקביל לפתח שנתי מערכות מתחרות: מערכת R שפותחה ב-IBM ומע' Ingers בברקלי. בהמשך המחקרים הבאים בססו את שפת SQL כשפה לאחזור מידע בבסיסי נתונים יחסיים. ובשנות ה-1980, פותחו מערכות לניהול מסד נתונים טבלאי.

### תיאור מודל נתונים יחסי

מודל יחסי לניהול בסיס נתונים הוא מודל נתונים המבוסס על תורת הקבוצות ולוגיקה של פרדיקטים. המודל היחסי היה המודל הוגדר כמודל פורמלי המבוסס על ההנחה שכל הנתונים מיוצגים כיחסים מתמטיים, כלומר כתת קבוצות שנתן להפעיל עליהן פעולות ויחסים כמו איחוד, חיתוך וכדומה. במודל מתמטי חישוב של ביטוי כזה יכול להניב אמת או שקר. במודל יחסי המתכנן יכול ליצור מודל לוגי למידע. מודל לוגי זה יכול להיות מתורגם בתהליך שנקרא נרמול לבסיס נתונים. בסיס הנתונים נבנה על מודל יחסי טהור היה מנורמל. כדי לאחזר מידע, משתמשים בשפות לאחזור נתונים דוגמת SQL.

מבנה הנתונים הבסיסי הוא יחס המתורגם לטבלאות, כאשר כל טבלה מכילה מידע על ישות מסויימת (לדוגמה, לקוחות במערכת בנקאית). ולכל רשומה בטבלה ראשית יש שדה שם, גיל ועיר מגורים. שורה

בטבלה נקראת גם tuple (קבוצה של n אלמנטים).

שתי

NAME	AGE	TOWN
Smith	29	
Jones	27	
Brooke	31	
Wilson	23	
Lomu	22	

Tuples or Rows

Attributes or columns

כל שורה היא רשומה מזהה של ישות אחת ואין שורות זהות באותה טבלה. עמודות מתארות תכונות של ישות. סדר השורות או עמודות אינו חשוב בטבלה.

לכל שורה (רשומה) יש **מפתח ראשי** (primary key), המשמש לזיהוי חד-משמעי שלה. מפתחות משניים משמשים לאיתור מהיר של רשומות

Fig 1

ולקישור בין רשומות בטבלאות שונות. ברשומת האזרח, מספר הזהות משמש כמפתח ראשי, ושם המשפחה והשם הפרטי משמשים כמפתח משני. מפתח המורכב מכמה שדות קרוי **מפתח מורכב**.

על יחסים ניתן להפעיל פעולות של אלגברה של יחסים כמו איחוד, חיתוך צירוף ועוד. לדוגמה, נתונות שתי טבלאות A ו-B:

Table A		Table B	
A	49	D	64
B	54	A	49

אם נבצע פעולת איחוד וחיתוך נקבל:

<p><b>Union</b> <math>(A \cup B):</math></p>	<p>תוצאת הפעולה איחוד של שני יחסים היא, יחס שלישי המכיל את כל השורות שבטבלה A וכל השורות בטבלה B, זו פעולה שקולה לפעולה הלוגית OR.</p>	<table border="1" style="margin: auto;"> <tbody> <tr> <td>A</td> <td>49</td> </tr> <tr> <td>B</td> <td>54</td> </tr> <tr> <td>D</td> <td>64</td> </tr> </tbody> </table>	A	49	B	54	D	64
A	49							
B	54							
D	64							
<p><b>Intersection</b> <math>(A \cap B):</math></p>	<p>תוצאת הפעולה חיתוך של שני יחסים היא, יחס שלישי המכיל את כל השורות שבטבלה A והן מופיעות גם בטבלה B, זו פעולה שקולה לפעולה הלוגית AND.</p>	<table border="1" style="margin: auto;"> <tbody> <tr> <td>A</td> <td>49</td> </tr> </tbody> </table>	A	49				
A	49							

כך במודל היחסי שימוש ביחסים ואלגברה של יחסים, מאפשר לבצע עיבוד על קבוצה של נתונים (ולא ברמה של רשומה שאפיינה את המודל ברשתי וההיררכי).

### מערכות לניהול בסיסי נתונים – DBMS (Database Management System)

מערכת לניהול בסיסי נתונים היא מערכת הכוללת אוסף של תוכניות המאפשרות למשתמשים ליצור ולנהל בסיסי נתונים כחלק של מערכת המידע. DBMS מסייעת להגדיר, להקים ולהפעיל בסיסי נתונים לצורך יישומים שונים, והיא גם כוללת שפה לאחזור מידע. ב-DBMS יש הפרדה ברורה בין נתונים ותוכניות. מבנה נתונים נשמר בקטלוג מתאין והוא יכול לשמש תוכניות רבות. במערכת יש שלוש רמות הפשטה (ABSTRUCTION):

- **רמה תפיסתית** – מודל המשתמש במושגים מעולמו של המשתמש ולא במונחי מחשב.
- **רמה פיסית** – מודל המתאר את אופן אחסון הנתונים במושגים פסיים של המחשב.
- **רמה לוגית** – מודל המתאר את בסיס הנתונים מנקודת המבט של המשתמש, אך עם זאת קרוב למבנה הראלי של בסיס הנתונים.

### המודל התפיסתי

מתאר את המערכת במונחים מובנים למשתמש:

- **ישות** – (ENTITY) אוביקט או עצם מהעולם הממשי.
- **תכונה** – (ATTRIBUTE) מאפיין של ישות (אם הישות היא לדוגמה סטודנט, אזי מס' זיהוי יכול להיות תכונה של הישות הזו).
- **קשר** – RELATIONSHIP יחסי גומלין בין שתי ישויות או בין ישות לעצמה.

- מודל ישויות קשרים (ER) – תרשים המתאר את מערכת הקשרים בין ישויות.
- זהו אחד המודלים התפיסתיים הנפוצים.

#### המודל הפיסי

- מתאר את אופן אחסון הנתונים במערכת המידע.
- המודל מכיל הגדרות של שדות הנתונים, מבנה הרשומות, שיטת ארגון הקבצים, שיטות החיפוש והגישה לנתונים, כולל אינדקסים.
- מרכיבים אלה הם שקופים למשתמש ומטופלים על ידי ה-DBA.

#### המודל הלוגי

- מתאר את הלוגיקה של ארגון הנתונים ואת היישום המעשי, תוך שהוא מתעלם מרוב המרכיבים הפיסיים. נזכיר כאן שלושה מודלים לוגיים עיקריים: היררכי, רשתי ויחסי.
- כל המודלים משתמשים במונחים מעולם המחשב: קבצים רשומות, שדות וכדומה.
- כל המודלים משתמשים בשפת תכנות מיוחדת שמשמשת להגדרת מבנה הנתונים.
- כמו כן המערכת מכילה שפה להגדרת נתונים - שפת ה-DDL (DATA DEFINITION LANGUAGE) ושפה לאחזור נתונים דוגמת SQL.

### טד קוד (Edger Frank Codd)

Edger Frank Codd שכונה בשם "Ted" נולד ב-1923 ומת ב-2003 בגיל 79. Codd פיתח את המודל היחסי לניהול בסיס נתונים ועל כך הוא קיבל ב-1981 את פרס טיורינג. מודל זה הווה גישה חדשה ומהפכנית לארגון נתונים, גישה כל כך פשוטה והגיונית שהפכה להיות הבסיס לניהול מערכות מידע. הוא נולד באנגליה בן לאב סוחר בעורות ואם מורה. הוא למד מתמטיקה וכימיה באוקספורד והיה קפיטן בחיל האוויר המלכותי במלחמת העולם השנייה. אחרי המלחמה עבר לארה"ב ולמד מתמטיקה באוניברסיטה עד 1949 כשהצטרף ל-IBM. הוא קיבל פרסים רבים וביניהם פרס טיורינג 1981 על פיתוח מודל הנתונים היחסי. הוא פרש מ-IBM ב-1984 והקים שתי חברות שסיפקו ייעוץ והרצאות ברחבי העולם לספקים ומשתמשי מערכות לניהול מסדי נתונים.

#### מקורות

1. A Relational Model of Data for Large Shared Data Banks  
*E. F. Codd*, Reprinted from *Communications of the ACM*, Vol. 13, No. 6, June 1970, pp. 377-387. Copyright © 1970, Association for Computing Machinery, Inc.
2. <http://wwbdb.web.cern.ch/wwbdb/aboutdbs/history/industry.html>
3. [http://www.pathology.ufl.edu/~computer/training/rdb\\_course/intro](http://www.pathology.ufl.edu/~computer/training/rdb_course/intro)