

מה ההבדלים בין: INNER JOIN ו- LEFT OUTER, RIGHT OUTER FULL JOIN ?

בעיקרון משתמשים בסוגים אחרים של JOIN (אחרים מ- INNER) כאשר יש מקרים שלא בטוחים באמינות הנתונים או שהנתונים באים ממקורות שונים ורוצים לתאם ביניהם.

דוגמה לשימוש:

נתונות הטבלאות הבאות:

Employee		Location	
EmpID	EmpName	EmpID	EmpLoc
13	Jason	13	San Jose
8	Alex	8	Los Angeles
3	Ram	3	Pune, India
17	Babu	17	Chennai, India
25	Johnson	39	Bangalore, India

שימו לב שבטבלה Employee ישנו עובד בשם Johnson שאיננו נמצא בטבלת המקומות וכן שישנו עובד בעל מספר זיהוי 39 שאינו נמצא בטבלת העובדים.

אם נכתוב Join פנימי (INNER JOIN) על מספר זיהוי עובד, יושמטו עובדים 25 ו-39. אבל אנחנו מעוניינים להתאים בין שמות העובדים ומיקומם, ורוצים לדעת גם אם עובד לא ידוע מיקומו. הטכניקה תהיה להשתמש במשהו שניקרא: LEFT JOIN או LEFT OUTER JOIN

לדוגמה:

```
select * from employee left outer join location
on employee.empID = location.empID;
```

אפשר גם להשמיט את מילת המפתח: OUTER ולכתוב:

```
select * from employee left join location
on employee.empID = location.empID;
```

התוצאה תהיה:

Employee.EmpID	Employee.EmpName	Location.EmpID	Location.EmpLoc
13	Jason	13	San Jose
8	Alex	8	Los Angeles
3	Ram	3	Pune, India
17	Babu	17	Chennai, India
25	Johnson	NULL	NULL

דבר דומה יקרה אם ניכתוב:

```
select * from employee right outer join location
on employee.empID = location.empID;
```

Employee.EmpID	Employee.EmpName	Location.EmpID	Location.EmpLoc
13	Jason	13	San Jose
8	Alex	8	Los Angeles
3	Ram	3	Pune, India
17	Babu	17	Chennai, India
NULL	NULL	39	Bangalore, India

תנחשו מה יקרה אם נכתוב:

```
select * from employee full outer join location
on employee.empID = location.empID;
```

המשך הדיון על הצפנה

מה יקרה אם אמזון תייצר את ההודעה וההצפנה ותשלח אותם לקרול על מנת לזהות את עצמה? $(x-y)$ אם מאלורי האזין וראה את x ואת y , הוא גם יכול לשמור אותם בצד ויום אחד להתחזות בתור אמזון ולשלוח אותם ערכים של x ו- y לבוב.

איך נפתור את הבעיה של הזיהוי (authentication)? כאן מגיע התפקיד של פונקציות האש חד-כיווניות.

$f(x)=y$

מהי פונקצית האש? נדגים זאת בעזרת משהו שימושי. נתונה לנו תוכנית בת 20,000 שורות בשפת C# ועלינו לספור כמה פעמים מופיעה כל מילת מפתח בתוכנית ולהדפיס בסדר עולה את המופעים. ונניח שאין לנו מחלקה בשם Dictionary למי שמכיר - לעבוד עם מערך בלבד.

ניכתוב פונקציית האש שהופכת כל מילת מפתח למספר בין 0 ל 50 וכך עם כל מופע של מילת מפתח x נבצע השמה: $arr[hash(x)] += 1$;

כלומר פונקציית האש תפקידה ליצור איזשהו קוד שיהיה עד כמה שאפשר ייחודי בטווח עבור כל ערך מערכי התחום האפשריים. לעיתים קוראים לה פונקציה ממפה.

כאשר קל לחשב את: $f(x)$, אבל מאד מאד קשה בהינתן y למצוא את x . ניקרא ל x בשם message ול- y ניקרא digest.

ה- one way hash גם יכולה לשמש כטביעת אצבע fingerprint. דוגמאות פופולאריות למשל: MD5 או SHA-1, יש גם SHA-512

האם פונקציה של lossless compression יכולה לשמש כפונקציית האש? לא - כי הקובץ ניתן לשיחזור. (uncompress). גם פונקציית checksum אינה יכולה לשמש (כמו בדיקה שאין שגיאות בהודעה), כי אפשר למצוא מספר הודעות בעלות אותו checksum

הצורה שהפונקציה עובדת, והיא בד"כ עם מחרוזות ביטים של 128, 160 או 256 או יותר ביטים. הפונקציה מערבבת את הביטים של ההודעה המקורית בצורות שונות ולבסוף לוקחת נניח 128 ביטים שמאליים אחרי הערבוביה. הרעיון הוא שאי אפשר לעשות reverse לפונקציית הערבוב, כלומר מהתוצאה לא ניתן לנחש את ההודעה. התוצאה של ההאש $h(x)$ ניקראת החתימה (signature)

שיעורי בית

כתובי את השאלות הבאות:

1. רשימה של כל המחברים שלהם יותר מספר אחד בטבלת הספרים.
2. רשימה של כל הסופרים שלפחות אחד מספריהם זמין.