

No Sql Databases

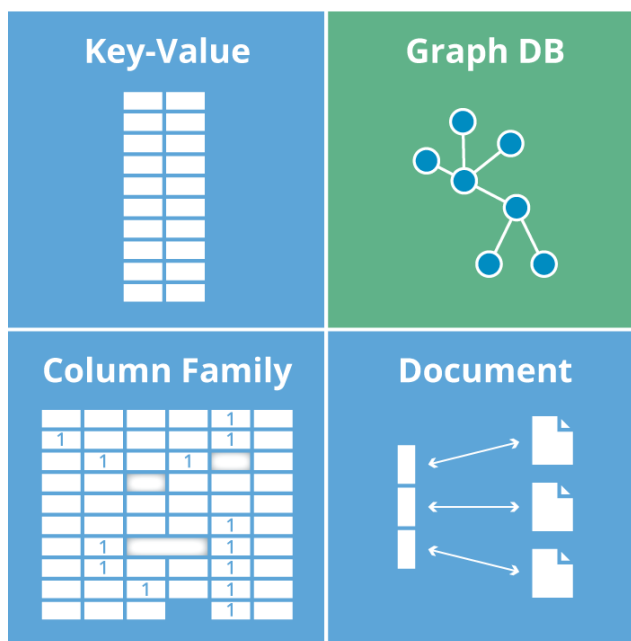
דן ארז

מהם No Sql Databases ולמה השם מטעה? אם נביט על ההיסטוריה של מסדי הנתונים, אז בעשורים האחרונים מסדי הנתונים הרלציוניים שלטו בכיפה. תמיד היו גם אחרים (למשל ObjectDB), אבל Relational DBs היו 'הזרם המרכזי'. בשנים האחרונות קמו מסדי נתונים שאינם רלציוניים, והם נקראים בשם הכללי NoSql Databases. הכוונה ב No היא ל Not Only, ולא ל 'לא sql' ולרבים מהם יש מימשיק .sql

למה בכלל צריך אותם? ובכן, הם שימושיים בכמה מקרים:

- מבנה עמודות גמיש/לא ידוע
- עבודה במודל שאינו רלציוני (למשל key-val או graph)
- בעיות scalability
- בעיות ביצועים (Too many joins?)
- אי התאמה לפורמטים 'מודרניים' (כמו json)
- מערכות מורכבות – אין db יחיד שעונה על כל הצרכים (יש חפיפה, אגב!)
- מחיר (רובם open source)

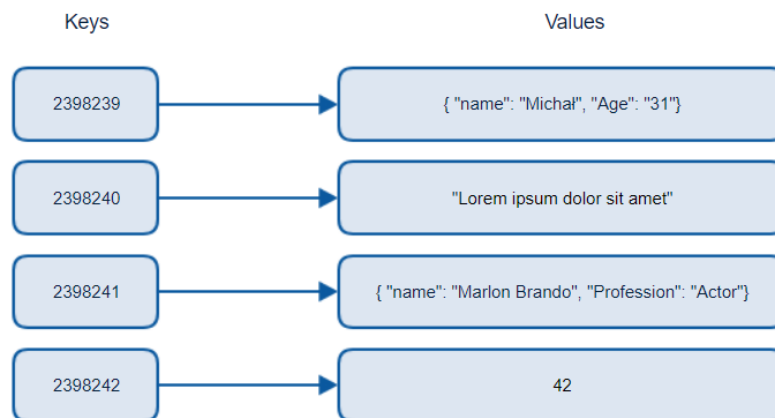
מסדי נתונים ממשפחת NoSql מתחלקים לכמה סוגים:



Key-Value DB

בגדול מדובר על מילון חכם או טבלת Hash, שמחזיק מיפוי בין Keys ל Values. מאפיינים:

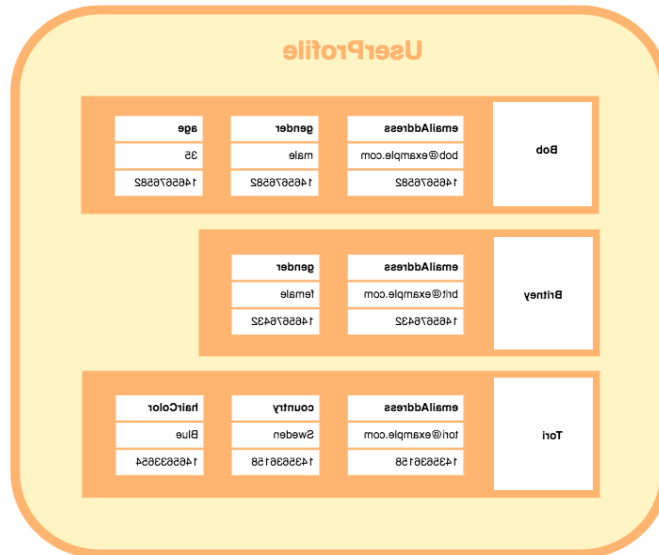
- שליפות לפי Keys
- ללא Joins
- דגש על מהירות
- לדוגמא: Redis, Riak, Amazon DynamoDB



Column DB

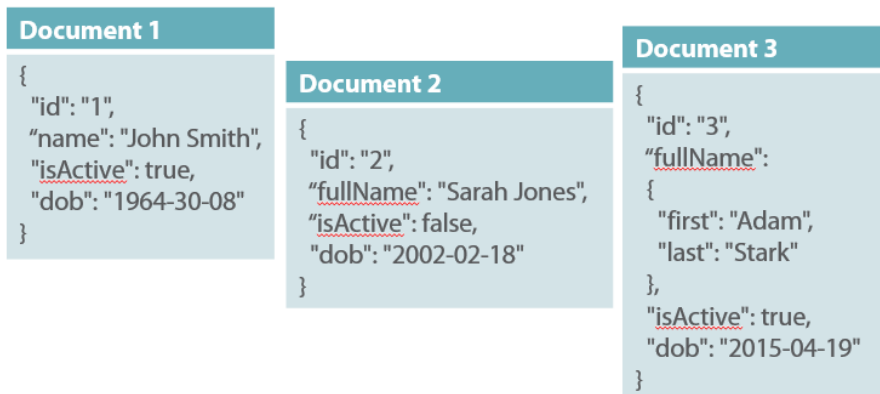
Column DB מקבץ נתונים לתוך עמודות בצורה גמישה. מאפיינים:

- עמודות במקום שורות (חיפוש מהיר)
- מבנה 'מוכר' – עמודות מקובצות ל'קבוצת עמודות'
- שמירה על גמישות מבנית
- שליפת מיליון כותרות משורות זה ארוך, אבל בעמודות זו גישה אחת לדיסק...
- Big Table
- למשל: Cassandra, HBase



Document DB

- התייחסות למידע כאל 'מסמך' (בד"כ json)
- יש מבניות מסויימת
- אינדוקס למטרות חיפוש
- Scalability in mind
- לדוגמה: MongoDB, CouchDB



Graph DB

- כאן המידע מסודר בצמתים ובקשרים ביניהם, תוך מעבר וחיפוש קל בגרף המתקבל.
- ריבוי קשרים (יותר מדי Joins)
 - שליפת תת גרפים
 - שפות תומכות (למשל cypher)
 - לדוגמה: Neo4J

