

הכשרת DBA

ליאור פולק

20 באפריל 2020

1 מקורות

1. MySQL
2. MongoDB
3. DB Normalization
4. Stanford DB class
5. Database Design Course

2 נושאים חשובים

1. Relational Algebra
2. BCNF (Functional dependencies: If $A \rightarrow B$, then A is a key)
3. 4NF (Multivalued dependencies: If $A \twoheadrightarrow B$, then A is a key)

3 Relational design by decomposition

1. Write "mega" relations + properties of the data
2. System decomposes based on properties
3. Final set of relations satisfies normal forms (no anomalies, no lost information): BCNF, 4NF

3.1 BCNF Normalization Algorithm

tl;dr: Separate tables until every functional dependency is a key, and only it given a table.

In every step we split by the key + value, and key + everything else.

3.2 4NF Normalization Algorithm

tl;dr: Separate tables until every functional dependency is a key, and only it given a table.

In every step we calculate functional dependencies and multivalued dependencies, and then split by the key + value, and key + everything else.¹

3.3 Shortcomings

1. Sometimes you may want to stop the algorithm before its completion, because the final tables may not contain the original dependencies without a join.
2. Sometimes you need redundancy - this is for computational reasons, when a set of values is commonly retrieved together.
3. Avoid over decomposition - tables that are too small.

¹Note that $4NF \Rightarrow BCNF$ since $A \rightarrow B \Rightarrow A \twoheadrightarrow B$.

4 Indices

Indexes are the difference between full table scan and immediate location of data.

Two types:

1. Tree types: used for $=, \leq, \geq$
2. Hash tables: used for $=$ only

4.1 Downsides

1. Index maintenance time (updated upon every update)
2. Index creation time
3. Extra space

5 Transaction

Transaction is a sequence of one or more SQL operations treated as a unit:

1. Transactions appear to run in isolation
2. Upon failure, each transaction's changes are reflected either entirely or not at all

ACID: Atomicity, Consistency, Isolation, Durability

1. Read Uncommitted- Dirty reads from middle of transaction, when performance matters and you only need approximate values.
2. Read Committed- Only read values that were committed to DB, still doesn't guarantee consistency (a transaction might end and we could read both of its sides)- default in MySQL
3. Repeatable Read- No dirty read, and multiple reads are consistent. A relation can change- phantom tuples can appear.
4. Serializable- guarantees consistency but slow- Standard

Read only mechanism allows the DB to adjust to a transaction that only reads.

6 NoSQL

A new paradigm for DB, stands for not only traditional relational DBMS.

Better handling of massive amounts of data, and efficiently.

Usually we don't require consistency, and we do graph operations.