

CS 338 course note

Chenxuan Wei

Sep 2022

Contents

| | | |
|----------|--|----------|
| 1 | Introduction to database | 3 |
| 2 | Relational | 4 |
| 3 | Relational algebra and calculus | 5 |

1 Introduction to database

1. Terms
 - Data redundancy: presence of duplicate data in multiple data files
 - Data inconsistency: the same attribute may have different values
2. Database
 - a collection of related information stored in a structured form
3. DBMS:
 - a collection of programs that manipulate a database
4. Data Model
 - Relational Model
 - Object-oriented model
 - semi-structured data model
 - network model
 - Hierarchical model
5. Schema
 - Physical schema: database at physical level
 - logical schema: database at logical schema
 - External schema: database at external schema

2 Relational

1. Terms

- attribute: each column with in a table
- domain: all possible value of a attribute
- Primary key: a attribute in a row that must be unique in a table
- Tuple: rows
- Schema of a relation: definiton of a table
- a instance: table content

2. Integrity Constaints

is a condition that must be true for any instance of the database

Domain constrain: must satisifeid domain

Primary key constraints: each relation must have a primary key, and they must be unique

Foreign key: set of filed in one relation used to refert to a tuple in another relation

3 Relational algebra and calculus

1. Relational Query language
A major strength of the relational model: supports simple, powerful querying of data
2. Relational algebra
Result of a retrieval is new relation
sequence of relational algebra operations forms a relational algebra expression
3. Operations
 - selection (σ): select a subset of rows from relation
 - projection(π) deletes unwanted columns from relation
 - cross-product(\times) allows us combine 2 relations
 - Set-difference ($-$) tuples in relation1 but not in 2
 - Union(\cup) tuple in both 1 and 2

Format: (operation)_{boolean} (relation)
4. Boolean
used to show true value
5. Assignment operation
< – allowed to assign variable
6. Union compatible
if 2 relations have the same degree and all attributes are defined on same domains
7. Foreign key
Assume R1(ABC), R2(EFG) there is a FK: R1.A references R2.G
the value of R1.A must be
Null or unique in R2
however, R2.G does not need to be PK
8. Rename operation (useless)
format: $\rho_{(relation)}(relation)$ or $\rho_{(col,col)}(relation)$
the first one rename relation, but the second one only rename column
9. Join operation
symbol: \bowtie
a combination of cross product and selection, notice must have different attributes name
The following are the same:
 - $e < -R1 \bowtie R2$
result $< -\sigma_{bool}(e)$

- $R1 \text{ (join)}_{bool}(R2)$
10. Natural join operation
 $\text{result} < - R1 * R2$
 Assume $R(ABC), S(AD), R * S \rightarrow (ABCD)$
 will auto=same attributes, and combine attributes, also allowed same attribute name
 11. Division Operation
 Assume $R1(r1_i), R2(r2_i), R1 \div R2 =$
 $(r1_i)$ such that $r1_i \notin R2$ and all removed $r1_i$ appear in every $R2$ tuple in $R1$
 12. Aggreation:
 $G_i g_{f_i(A_i)}(E)$, allowed optional As to change the name of function $F1$
 function includes
 - avg
 - min
 - max
 - sum
 - count