Data Management

Leonardo Maria Carrozzo

2022/2023

Contents

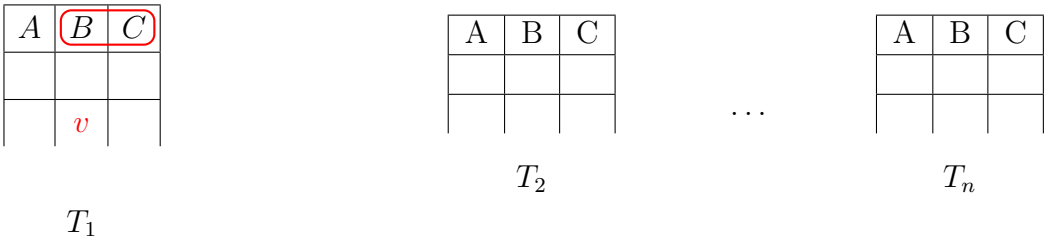
1 Introduction

The course is based on the following topics:

- **The structure of a Data Base Management System (DBMS)**: Realtional data and queries, Buffer manager;
- **Transaction management**: The concept of transaction, Concurrency management;
- **Crash management**: Classification of failures, Recovery;
- **Data Warehousing**: Data warehousing architectures and operators, Data warehousing design;
- **NoSQL databases**: Document-based databases (such as MongoDB), Graph databases OLAP vs OLTP (such as Neo4j);
- **Physical structures for data bases**: File organizations for data base management, Principles of physical database design;
- **Query processing**: Evaluation of realational algebra operatos, Fundamentals of query optimization;

1.1 The relational data model

A **database** in the Realtional Model is a **set of tables** (or **relations**). Each **table** is a **set of rows** (or **tuples**). Each one with the **same set of columns** (or **attributes**).



v is the value of the corresponding column and row. The attributes B and C form a **superkey**.

We have then:

- **Integrity constraint**: a rule at the level of the schema that all the rows must respect;
- **Superkey**: there cannot exist two or more rows that have the same value as the combination of multiple attributes;

- **Key**: attribute in a table;
- **Foreign key**: attributes in a table are a reference of another table;
- **Primary key**: special key that doesn't allow null values (a null value is a a special value that says that the value is missing).

A	B	C
		c_1
		c_2
		c_3
		c_3

T_1 Unordered set

D	E	F
c_0		
c_1		
c_2		
c_3		

T_2 Must not miss any key

A	B	C
a_1	b_1	c_1
a_1	b_2	c_2
a_2	b_2	c_3
$null$	b_2	c_3

T_1 Unordered set

D	E	F
c_0		
c_1		
c_2		
c_3		

T_2 Must not miss any key

We have a "no predicate" on the null value: never equal and never different, comparation is always false.

As we have said, the Relational Data Model uses the mathematical concept of a relation as the formalism for describing and representing data. A relation is a subset of a cartesian product of sets. A relation can be considered as a "table" with rows and columns.

Codd introduced two different query languages for the relational data model:

- **Relational Algebra**, which is a procedural language. It is an algebraic formalism in which queries are expressed by applying a sequence of operations to relations.
- **Relational Calculus**, which is a declarative language. It is a logical formalism in which queries are expressed as formulas of fist-order logic.

Codd's Theorem: Relational Algebra and Relational Calculus are essentially equivalent in terms of expressive power.

DBMSs are based on **SQL**, a hybrid of a procedural and a declarative language that combines features from both relational algebra and relational calculus.

1.2 Relational algebra