Database Management Systems

Part II: The Relational Model

Lecture 5

Relational Data Integrity: Candidate Keys and Related Matters

Contents

- Introduction
- Relational Data Integrity
- Candidate Keys
- Primary Keys
- Alternate Keys
- Foreign Keys
- Nulls

Introduction

- In a real-world application, a database contains thousands of tables which include thousands of records.
- In such cases, there is a challenge to uniquely identify a table record.
- To overcome this, key is used to identify any row of data in a table.
- DBMS keys allow to establish a relationship between the tables and identify the relation between them.

Introduction (Cont.)

- In DBMS, a key is an attribute or set of an attribute to identify a tuple (row) in a relation (table).
- By using key, we can uniquely identify a table record despite the challenges.
- Moreover, the relation between the two tables can be identified.
- Consequently, it can enforce the identity and integrity in the relationship.

Relational Data Integrity

- The relational model does include two general integrity features.
- These features are relevant to the integrity of every database, not just to some specific database.
- These two features are
  1. Candidate (and primary) keys, and
  2. Foreign keys.

Candidate Key

- Candidate key is a set of one or more than one columns (attributes) which uniquely identifies each record in a table.
- There must not be redundant values (repetition of cells) in selected attribute.
- The candidate key must contain unique values.
- Also, it must not contain null values.

Candidate Key (Cont.)

- Every relation does have at least one candidate key because relations do not contain duplicate tuples.
- That is, since tuples are unique, it follows that at least the combination of all attributes of the relation has the uniqueness property.
- Candidate keys are fundamental to the successful operation of a relational system.

Candidate Key (Cont.)

- For example, Roll No. is unique in STUDENT relation.
- This can be selected as a candidate key.
- If there exist the other unique attributes in the relation, we can then select more than one column as candidate key to uniquely identify a record.

Candidate Key (Cont.)

- In the given table, Student_ID, Roll_No. and Email are candidate keys which can uniquely identify the student record in the table.

Student

Student_ID	Roll_No.	Student_Name	Major	Email
1	IT-1	Mg Mg	IT	mgmg@gmail.com
2	IT-2	Ma Ma	IT	mama@gmail.com
3	IT-3	Ko Ko	IT	koko@gmail.com

Primary Key

- Primary key of a relation is just a unique identifier for that relation.
- Primary key is used to uniquely identify a record in relation.
- The primary keys are compulsory in every table.

Primary Key (Cont.)

- Primary keys are really only a special case of the more fundamental concept candidate key.
- The primary key cannot be duplicated.
- The primary key field cannot be null.
- The value in a primary key column can never be modified or updated if any foreign key refers to that primary key.

Primary Key (Cont.)

- The primary keys are having model stability, occurrence of minimum fields, being definitive and feature of accessibility.
- In the following example, Student_ID is a primary key.

Student

Student_ID	Roll_No.	Student_Name	Major	Email
1	IT-1	Mg Mg	IT	mgmg@gmail.com
2	IT-2	Ma Ma	IT	mama@gmail.com
3	IT-3	Ko Ko	IT	koko@gmail.com

Foreign Key

- Foreign key is a key of one table which points to the primary key in the second table.
- It has a relationship with the primary key in the second table.
- Foreign keys help us to maintain data integrity and also allows navigation between two different instances of an entity.
- Every relationship in the model needs to be supported by a foreign key.

Foreign Key (Cont.)

- In this example, we have the two tables, Student and Mark.
- Student_ID is primary key for Student relation and Roll_No. is primary key for Mark relation.
- However, there is no way to see who got which marks.

Student

Student_ID	Student_Name	Major	Email
1	Mg Mg	IT	mgmg@gmail.com
2	Ma Ma	IT	mama@gmail.com
3	Ko Ko	IT	koko@gmail.com

Mark

Roll_No.	Marks
IT-1	80
IT-2	85
IT-3	76

Foreign Key (Cont.)

- To create a relationship between the two tables, Student_ID column from the Student relation is added to the Mark relation to be used as the foreign key.
- This concept is also known as Referential Integrity.

Mark

Roll_No. (Primary Key)	Student_ID (Foreign Key)	Marks
IT-1	1	80
IT-2	2	85
IT-3	3	76

Other Keys

- There exist the other useful keys in database management system.
- They are
  - ✓ Super Key
  - ✓ Alternate Key

Other Keys (Cont.)

- A super key is a group of single or multiple keys which identifies rows in a table.
- Super key is a super set of candidate key.
- A super key may have additional attributes that are not needed for unique identification.

EmpSSN	EmpNum	Empname
9812345098	AB05	Shown
9876512345	AB06	Roslyn
199937890	AB07	James

- In the above-given example, EmpSSN and EmpNum name are super keys.

Other Keys (Cont.)

- All the keys which are not primary key are called alternate keys.
- In this table, Student_ID, Roll_No. and Email are qualified to become a primary key.
- Since Student_ID is selected as the primary key, Roll_No. and Email become the alternate keys.

Student

Student_ID	Roll_No.	Student_Name	Major	Email
1	IT-1	Mg Mg	IT	mgmg@gmail.com
2	IT-2	Ma Ma	IT	mama@gmail.com
3	IT-3	Ko Ko	IT	koko@gmail.com

Null

- A null value indicates an absent value that may exist but be unknown or that may not exist at all.
- Basically, nulls are intended as a basis for dealing with the problem of missing information.
- The idea is that if a given tuple has a null in a given attribute position, it means that the value of that attribute is missing for some reason for the tuple in question.

Null (Cont.)

- In general, there have been several meanings for NULL values, such as value unknown, value exists but is not available, or attribute does not apply to this tuple (also known as value undefined).
- Nulls are not the same as blank or zero.
- It is a special value that signifies that the value is unknown or does not exist.

Null (Cont.)

- Null values present special problems in relational operations, including arithmetic operations, comparison operations, and set operations.
- For example, the result of an arithmetic expression (+, −, ∗, or /) is null if any of the input values is null.

Summary

- Candidate keys are required to be unique and irreducible.
- Every relation has at least one candidate key.
- One candidate key is designated as the primary key and any other candidate keys were alternate keys.
- Null is a scheme based on default values.
- No component of the primary key of a base relation is allowed to accept nulls.
- The relation model has permitted nulls to appear in foreign key positions.

Next Lecture

Part II: The Relational Model

Relational Operators I

- Relational Algebra
- Traditional Set Operations
- Special Relational Operations
- Extend and Summarize
- Relational Comparisons

Textbook and References

Textbook

- C. J. Date, “An Introduction to Database Systems”, 6th Edition, 1994.

Additional References

- Abraham Silberschatz, Henry F. Korth, S. Sudarshan, “Database System Concepts”, 6th Edition, 2011.
- Ramez Elmasri, Shamkant B. Navathe, “Fundamentals of Database Systems”, 6th Edition, 2010.
- https://www.guru99.com
- https:// www.includehelp.com
- https:// www.tutorialspoint.com