

CS245: Databases

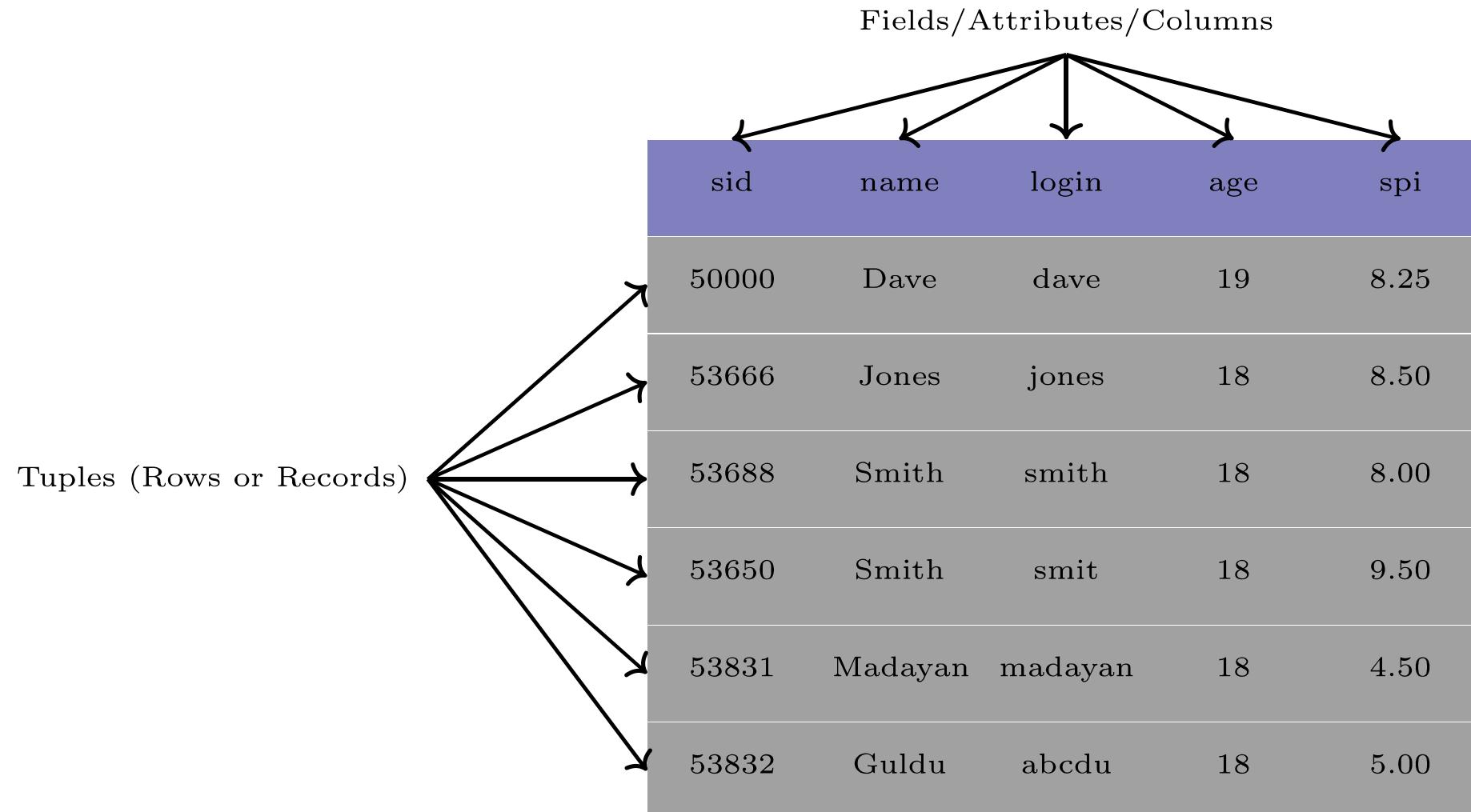
Introduction

Vijaya saradhi

Department of Computer Science and Engineering
Indian Institute of Technology Guwahati

Tables

Table Notations



Need for constraints on table

student				
sid	name	login	age	spi
190101000	Atul Kumar	atul	18	8.0
190101000	Atul Gupta	atul	18	8.2
190101000	Atul M	atul	18	8.2
190101000	Atul Gupta	atul	19	7.2

- Same roll number is assigned to several students
- Same login is assigned to several students
- It is not possible to distinguish between two Atul Gupta's (row 2 & 4)
- In case you have to update the spi of Atul Gupta which row will you update? 2 or 4?

Constraints on Tables

Not discussing all the constraints at present

- Primary key
- Uniqueness
- Not NULL
- DEFAULT
- (requires) Two or more tables - Foreign key

single column

- One column designated as primary key
- When primary key column has identical values then corresponding rows must have identical values OR
- Primary key values must be all distinct

single column

- One column designated as primary key
- When primary key column has identical values then corresponding rows must have identical values OR
- Primary key values must be all distinct

single column - example 01

sid is single column primary key

Violation of primary key constraint

student				
sid	name	login	age	spi
190101000	Atul Kumar	atul	18	8.0
190101001	Atul Gupta	atul	18	8.2
190101001	Atul M	atul	18	8.2
190101002	Atul Gupta	atul	19	7.2

Primary key - 01

single column

- One column designated as **primary key**
- When primary key column has **identical values** then corresponding rows must have **identical values** OR
- Primary key values must be all **distinct**

single column - example 02

sid is single column primary key

No violation of primary key constraint

However, database engines will not allow two identical values in primary key column

student				
sid	name	login	age	spi
190101000	Atul Kumar	atul	18	8.0
190101000	Atul Kumar	atul	18	8.0
190101001	Atul M	atul	18	8.2
190101002	Atul Gupta	atul	19	7.2

Two columns

- Two column combinely described as primary key
- When primary key **columns** has identical values then corresponding rows must have **identical values** OR
- Primary key values must be all **distinct**

Two columns

- Two column combinely described as primary key
- When primary key **columns** has identical values then corresponding rows must have **identical values** OR
- Primary key values must be all **distinct**

two columns - example 01

{**sid**, **cid**} together are primary key

Violation of primary key constraint

register		
sid	grade	cid
190101000	AB	CS101
190101000	BB	CS101
190109001	AA	CS101
190109001	BB	CS102

Primary key - 02

Two columns

- Two column combination described as primary key
- When primary key **columns** has identical values then corresponding rows must have **identical values** OR
- Primary key values must be all **distinct**

two column - example 02

{**sid**, **cid**} together are primary key

No violation of primary key constraint

However, database engines will not allow two identical values in primary key column

register		
sid	grade	cid
190101000	AB	CS101
190101000	AB	CS101
190109001	AA	CS101
190109001	BB	CS102

All columns

- All the columns combinedly described as primary key
- When primary key **columns** has identical values then corresponding rows must have **identical values** OR
- Primary key values must be all **distinct**

All columns

- All the columns combinedly described as primary key
- When primary key **columns** has identical values then corresponding rows must have **identical values** OR
- Primary key values must be all **distinct**

All columns - example 01

{**sid, year, cid**} together are primary key
Is this primary key constraint violation?

register		
sid	year	cid
190101000	2020	CS101
190101000	2020	CS101
190109001	2021	CS101
190109001	2022	CS102

All columns

- All the columns combinedly described as primary key
- When primary key **columns** has identical values then corresponding rows must have **identical values** OR
- Primary key values must be all **distinct**

All columns - example 02

{**sid, year, cid**} together are primary key
Is this primary key constraint violation?

register		
sid	year	cid
190101000	2020	CS101
190101000	2021	CS101
190109001	2021	CS101
190109001	2022	CS102

Why Primary?

In the student table example

- **sid** is a key.
- login also can be a key.
- No two students can have identical login values.
- We choose one of them to be the **primary key**
- All queries use to **sid** for convenience
- It is possible that queries may use login key instead of **primary key** to retrieve data

student				
sid	name	login	age	spi
190101001	Atul Kumar	atul	18	8.0
190101002	Atul Kumar	ak	18	8.0
190101003	Atul M	atulm	18	8.2
190101004	Atul Gupta	atulg	19	7.2

example - 01

- What is the spi of student with **sid** 190101001?
- What is the spi of student with **login** atul?
- Can you query: What is the spi of student with **name** “Atul Kumar”?
- Can you query: What is the spi of student with **age** 18?
- Last two queries are not erroneous. They result in retrieving multiple rows.

Description

- Exists purely to identify rows of a table (relation/entity)
- Do not imply any property of instances
- Example: Order number, product code, batch number, etc.

Details

IDs may be of three types

- System generated
- Administrator generated
- Externally defined identifiers

Examples

- Order numbers (no human intervention)
- Account numbers, RD number, FD number, mobile number, etc..
- Generated in **sequence** without any specific requirement of the sequence generation
- Can be numeric and non-numeric

Examples

- Only suitable for relatively **low-volume** entity classes
- Department codes, product codes, class room numbers, course codes etc
- Can be numeric or non-numeric
- Administrator have mechanism to create new identifiers

Examples

- Defined by external party
- Often by national or international standards authority
- Country codes (telephone numbers)
- Currency codes
- State codes
- Pin codes
- Codes externally defined but administrator generated for postal department

Role

- Used in many instances of operations
- Used as constraints
- Uniquely identifying rows of a table

Primary key - 05

Primary Key



sid	name	login	age	spi
50000	Dave	dave	19	8.25
53666	Jones	jones	18	8.50
53688	Smith	smith	18	8.00
53650	Smith	smit	18	9.50
53831	Madayan	madayan	18	4.50
53832	Guldu	abcd	18	5.00

Primary key - 06

Second row is legal when there is no constraint on login column

Primary Key



sid	name	login	age	spi
50000	Dave	dave	19	8.25
50001	Dave	dave	19	8.25
53666	Jones	jones	18	8.50
53688	Smith	smith	18	8.00
53650	Smith	smit	18	9.50
53831	Madayan	madayan	18	4.50
53832	Guldu	abcd	18	5.00

Primary key violation - 07

Cannot have two rows having identical sid values

Primary Key



sid	name	login	age	spi
50000	Dave	dave	19	8.25
53666	Jones	jones	18	8.50
53688	Smith	smith	18	8.00
53650	Smith	smit	18	9.50
53831	Madayan	madayan	18	4.50
53832	Guldu	abcd	18	5.00
53666	James	james	18	8.50

Primary key - 08

More than one column can participate in Primary Key

Primary Key		
cid	grade	sid
CS101	CC	53831
CS203	BB	53832
CS112	AB	53650
HS105	BB	53666

Relation between course table and student table

Primary Key		Primary Key	
cid	grade	sid	
CS101	CC	53831	
CS203	BB	53832	
CS112	AB	53650	
HS105	BB	53666	
HS106	AB	53666	

One student registering for two courses

Primary key - 10

Relation between course table and student table

Primary Key		Primary Key	
cid	grade	sid	
CS101	CC	53831	
CS203	BB	53832	
CS112	AB	53650	
HS105	BB	53666	
HS105	BB	53832	Two students registering for one course

Primary key - 11

Relationship between three tables

Register		
cid	grade	sid
CS101	CC	53831
CS203	BB	53832
CS112	AB	53650
HS105	BB	53666
HS105	BB	53832

Student		
sid	name	spi
50000	Dave	8.25
53666	Jones	8.50
53688	Smith	8.00
53650	Smith	9.50
53831	Madayan	4.50
53832	Guldu	5.00
53666	James	8.50

Course		
cid	cname	credits
CS101	C Programn	3-0-0-6
CS203	Algorithm	3-0-0-6
CS112	Web Programn	0-1-3-5
HS105	Economics	3-0-0-6
HS105	Political Sci.	3-0-0-6

Primary key - 12

Relationship between three tables

Grade		Student	Course
cid	grade	sid	cid
CS101	CC	53831	50000
CS203	BB	53832	53666
CS112	AB	53650	53688
HS105	BB	53666	53650
HS105	BB	53832	53831
			53832
			53666

Primary key - 13 violation

None of the columns of the primary key should have \perp values

Primary Key		
cid	grade	sid
CS101	CC	53831
CS203	BB	53832
CS112	AB	53650
HS105	BB	53666
\perp	BB	53832

Why NULL values become an issue

register		
cid	grade	sid
CS101	CC	53831
CS203	BB	53832
CS112	AB	53650
HS105	BB	53666
⊥	BB	⊥
⊥	BB	⊥

Cannot disting two \perp values. That is the test: $\perp == \perp$ will NOT evaluate to TRUE!

Primary key - 14 violation

None of the columns of the primary key should have \perp values

Primary Key		
cid	grade	sid
CS101	CC	53831
CS203	BB	53832
CS112	AB	53650
HS105	BB	53666
HS105	BB	\perp

Primary key - 15 more than one column

Can these three columns put together be primary key? - yes

Primary Key		
cid	grade	sid
CS101	CC	53831
CS203	BB	53832
CS112	AB	53650
HS105	BB	53666
HS106	BB	53666

One student registering for two courses;

Primary Key - 16 more than one column

Observe for change in meaning

Primary Key		
cid	grade	sid
CS101	CC	53831
CS203	BB	53832
CS112	AB	53650
HS105	BB	53666
HS106	BB	53666

- Allows two or more students to register for one course
- Allows one student to register for two or more courses

allow one student to register more than one course

Primary key - 17 more than one column

Meaning of two columns to be the primary key

Primary Key		
cid	grade	sid
CS101	CC	53831
CS203	BB	53832
CS112	AB	53650
HS105	BB	53666
HS105	BB	53667

- Allows two or more students to register for one course
- Allows one student to register for two or more courses

allow one course to register more than one student

Primary key - 18 more than one column

Observe for change in meaning

Primary Key		
cid	grade	sid
CS101	CC	53831
CS203	BB	53832
CS112	AB	53650
HS105	BB	53666
HS106	BB	53666

Cannot allow one student to register for two different courses

Primary key - 19 more than one column

Observe for change in meaning

Primary Key

cid	grade	sid
CS101	CC	53831
CS203	BB	53832
CS112	AB	53650
HS105	BB	53666
HS105	BB	53667

Cannot allow multiple students to register for one course

Uniqueness

Primary Key		Uniqueness		
↓	↓			
sid	name	login	age	spi
50000	Dave	dave	19	8.25
53666	Jones	jones	18	8.50
53688	Smith	smith	18	8.00
53650	Smith	smit	18	9.50
53831	Madayan	madayan	18	4.50
53832	Guldu	abcd	18	5.00

Uniqueness Violation

Cannot have two rows having identical login values

Primary Key		Uniqueness		
		↓	↓	
sid	name	login	age	spi
50000	Dave	dave	19	8.25
53666	Jones	jones	18	8.50
53688	Smith	smith	18	8.00
53650	Smith	smit	18	9.50
53831	Madayan	madayan	18	4.50
53832	Guldu	abcd	18	5.00
53835	Dave D	dave	19	7.00

Not NULL - has implicit meaning - 01

Primary Key	Uniqueness		Not Null	
↓	↓	↓	↓	↓
sid	name	login	age	spi
50000	Dave	dave	19	8.25
53666	Jones	jones	18	8.50
53688	Smith	smith	18	8.00
53650	Smith	smit	18	9.50
53831	Madayan	madayan	18	4.50
53832	Guldu	abcd	18	5.00

Not NULL - has implicit meaning - 02

Cannot have a cell taking \perp (NULL) value

Last row is illegal

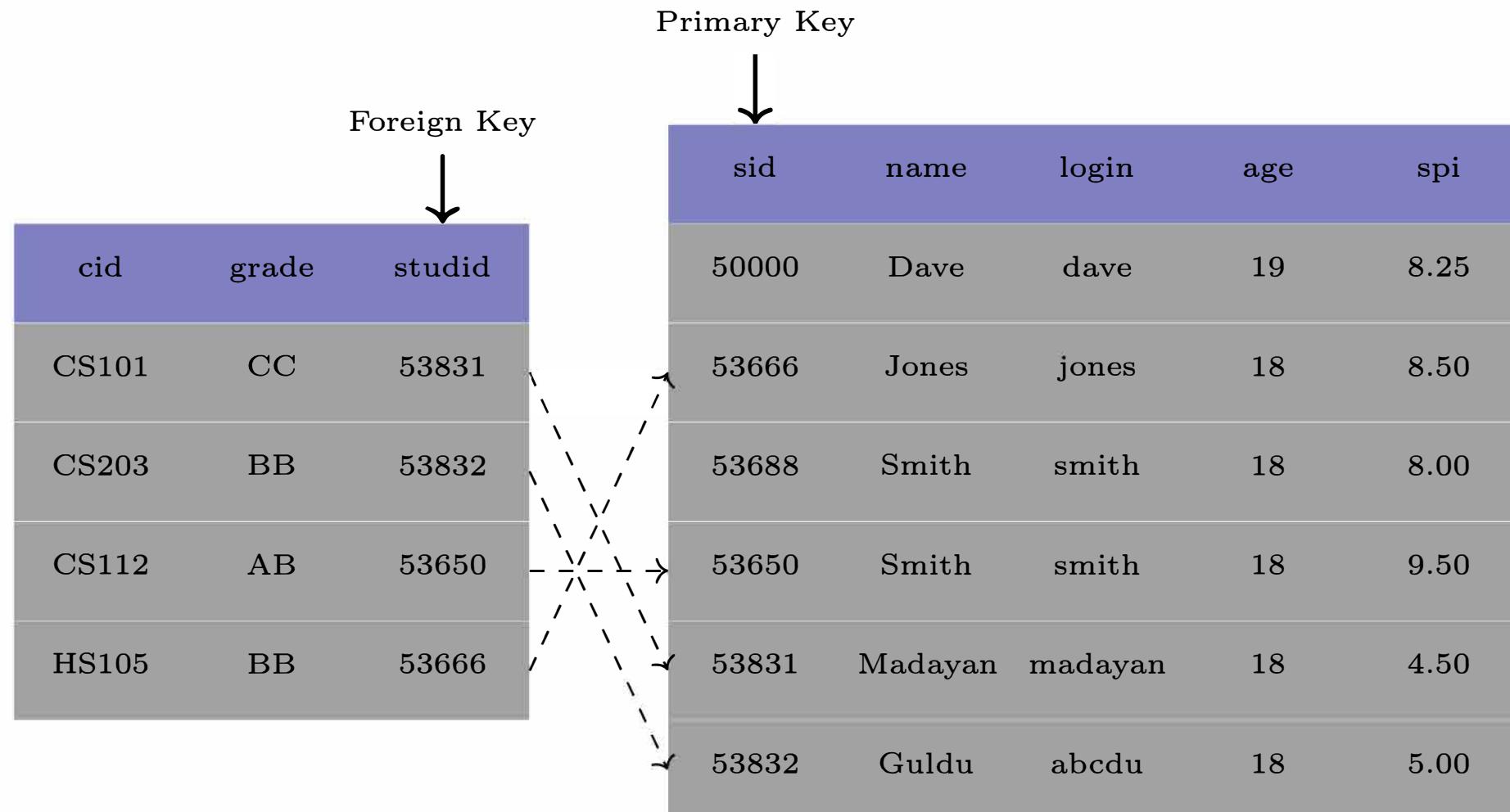
Primary Key		Uniqueness		Not Null
	↓	↓	↓	
sid	name	login	age	spi
50000	Dave	dave	19	8.25
53666	Jones	jones	18	8.50
53688	Smith	smith	18	8.00
53650	Smith	smit	18	9.50
53831	Madayan	madayan	18	4.50
53832	Guldu	abcd	18	5.00
53835	Atul	atulp	19	\perp

DEFAULT value

Primary Key		Uniqueness		Not Null
↓	↓	↓	↓	↓
sid	name	login	age	spi
50000	Dave	dave	19	8.25
53666	Jones	jones	18	8.50
53688	Smith	smith	18	8.00
53650	Smith	smit	18	9.50
53831	Madayan	madayan	18	4.50
53832	Guldu	abcd	18	5.00
53835	Atul	atulp	100	8.5

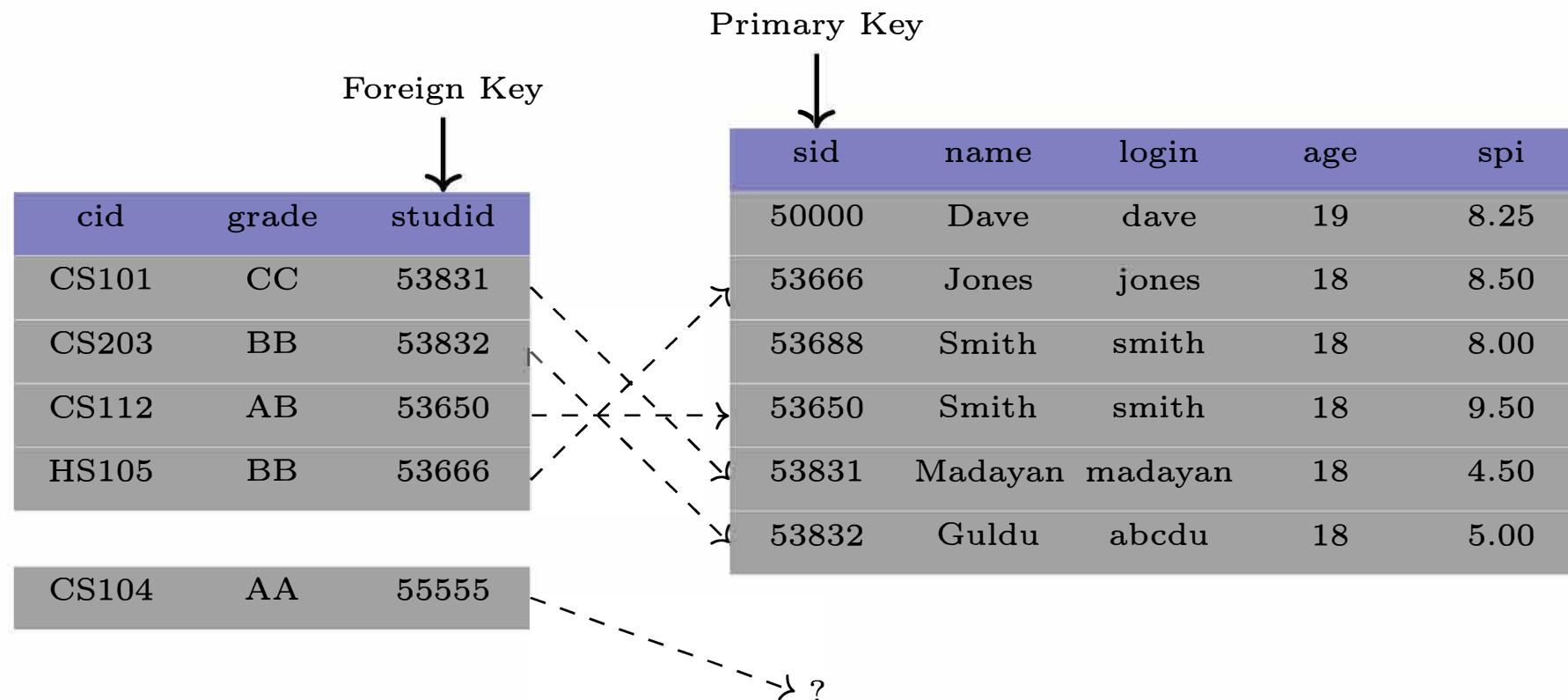
- Assume column `age` has DEFAULT value 100
- Assume you are inserting a new row
- You have specified values for `sid`, `name`, `login`, and `spi`
- Not specified any value for `age`
- The DEFAULT constraints is responsible to write value 100 to the `age` of the new row

Foreign Key

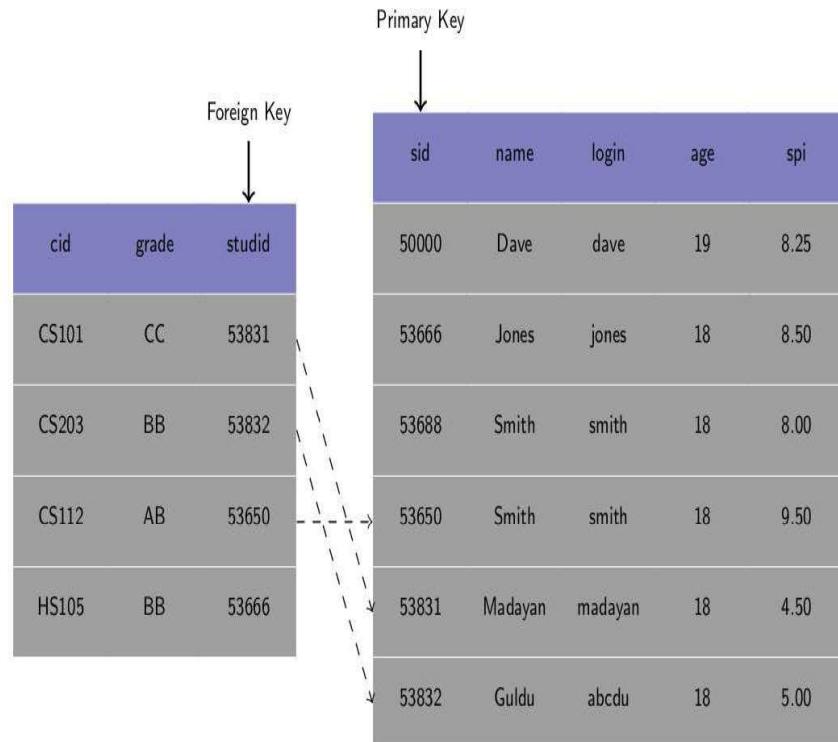


Foreign Key (Referential Integrity)

Cannot insert CS104 into LHS table as studid=55555 doesn't exist in RHS table sid column



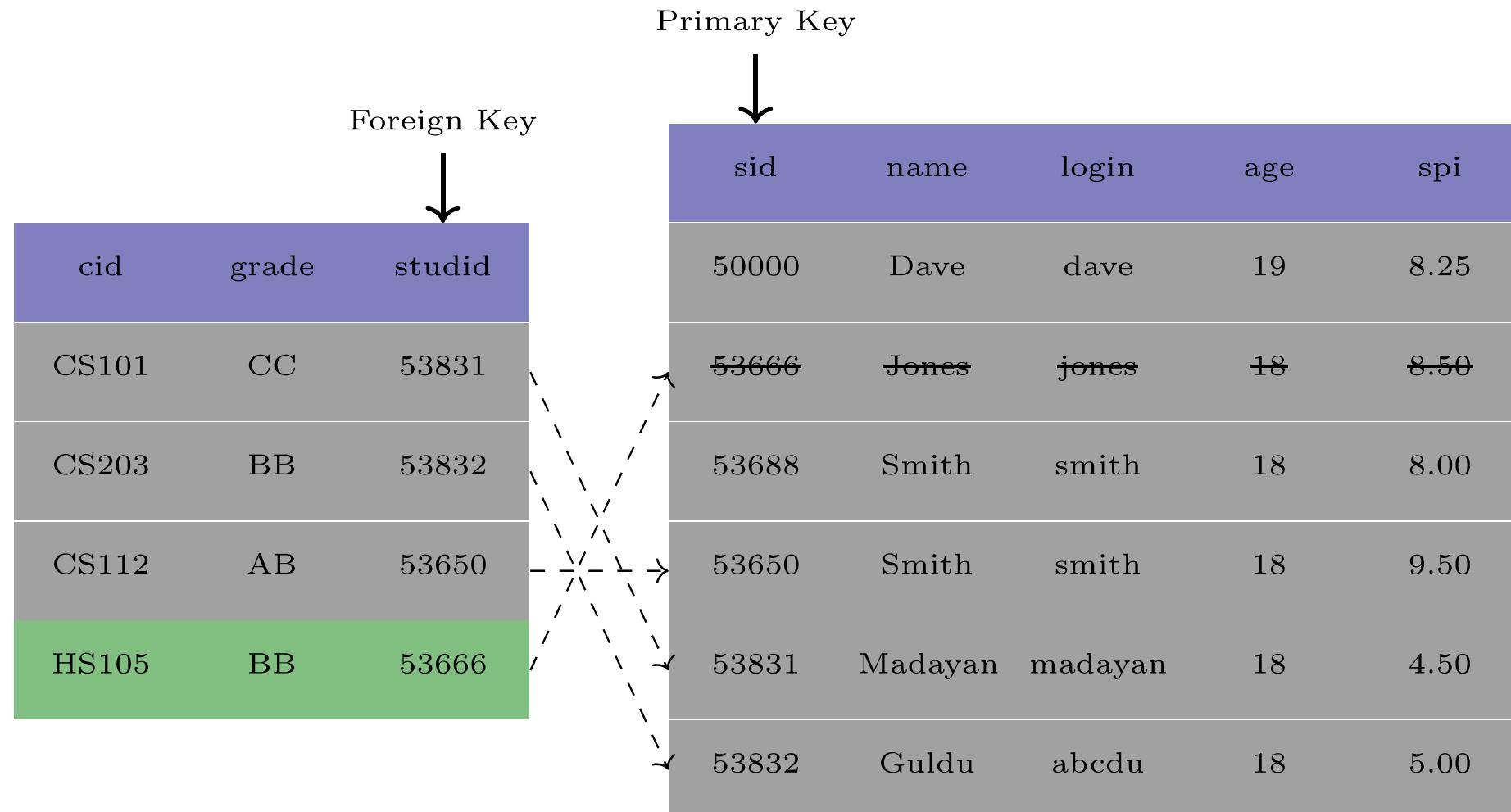
Foreign Key - Cases



Operations on LHS & RHS tables

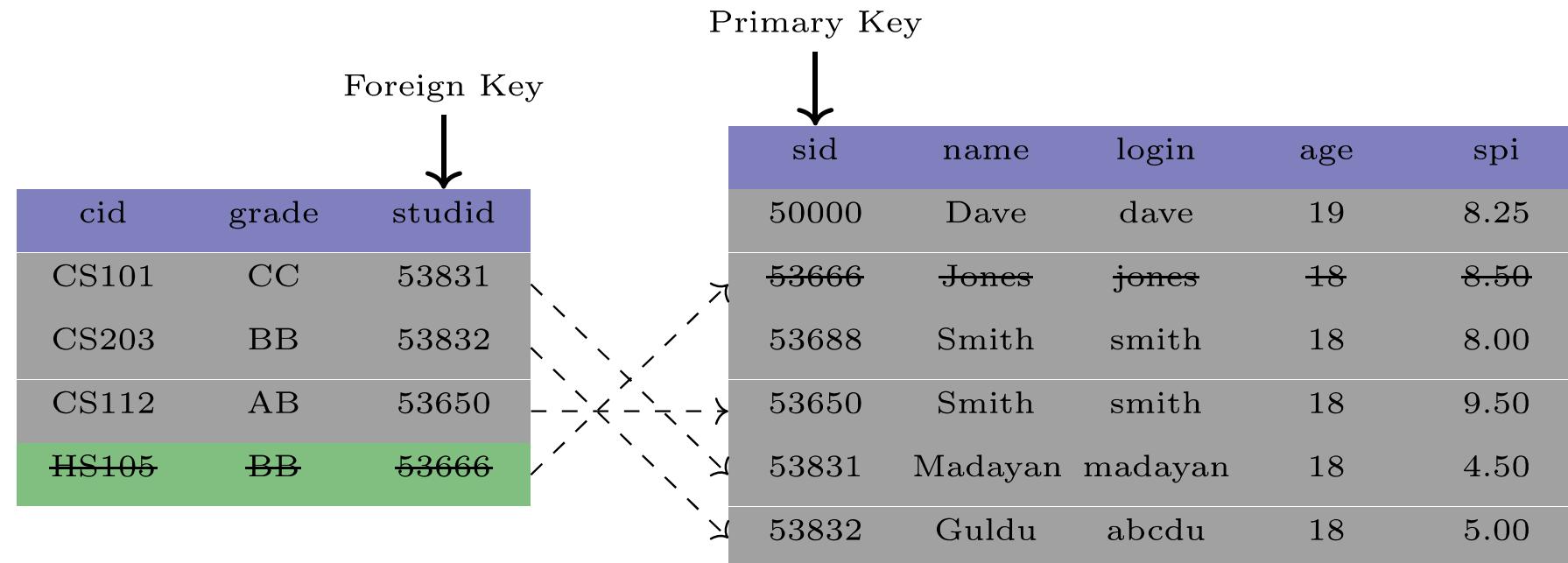
- 1 A row is deleted from RHS table
- 2 A row is updated in RHS table
- 3 A row is inserted into LHS table
- 4 A row is deleted from LHS table

Foreign Key - Row Deletion from RHS Table



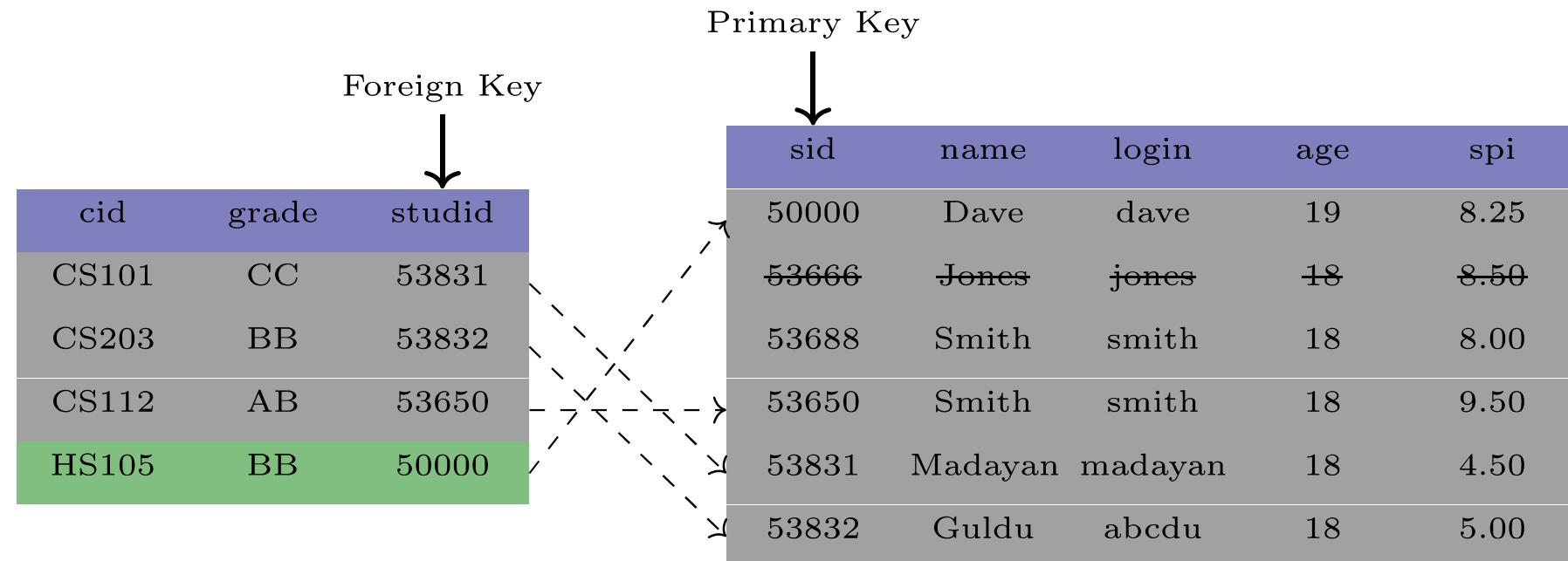
Foreign Key - Row Deletion from RHS Table - Action 01

Delete all rows in LHS table with studid=53666



Foreign Key - Row Deletion from RHS Table - Action 02

Write a DEFAULT values (say 50000) in all rows in LHS table with studid=53666

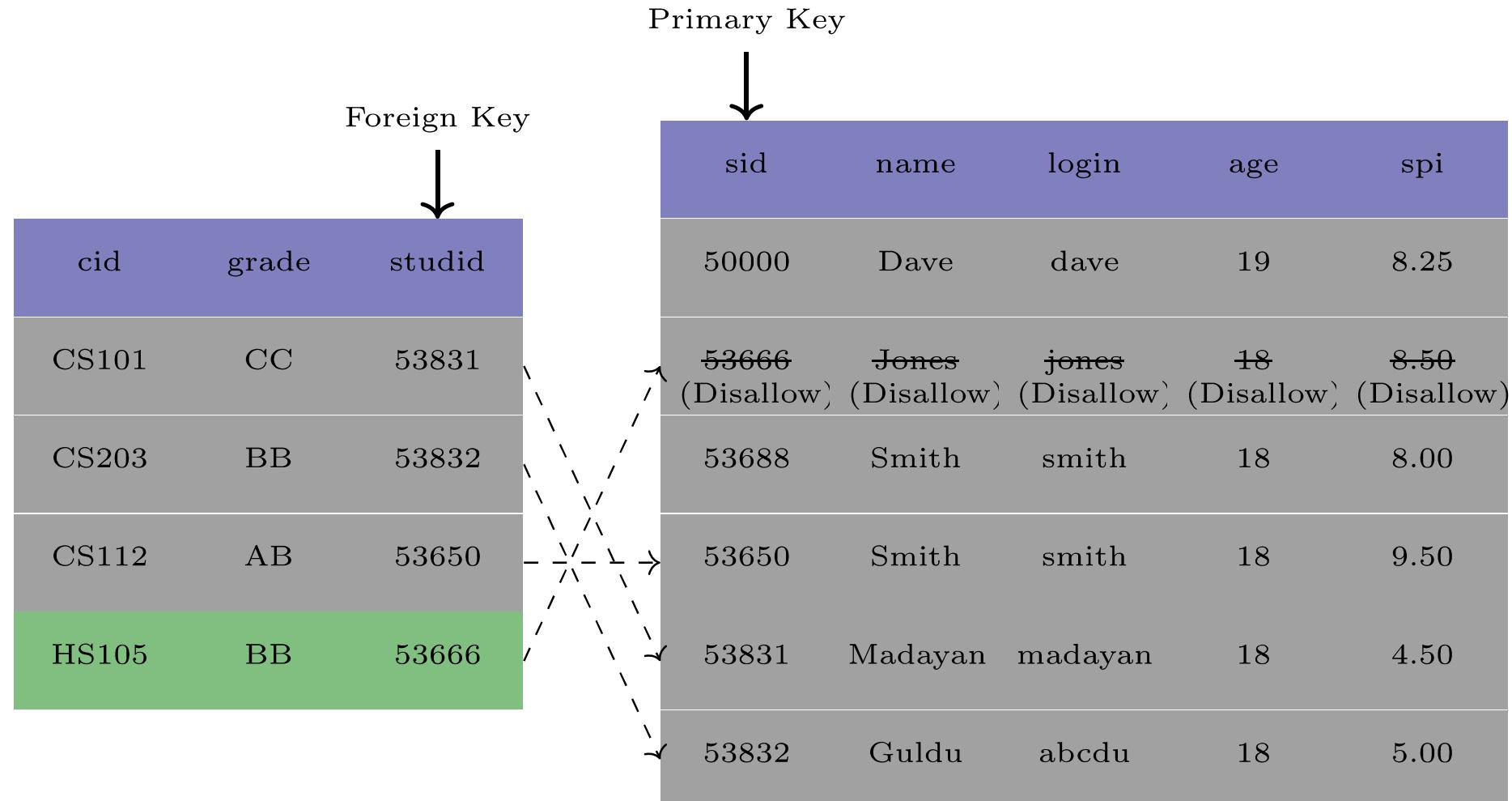


Foreign Key - Row Deletion from RHS Table - Action 03

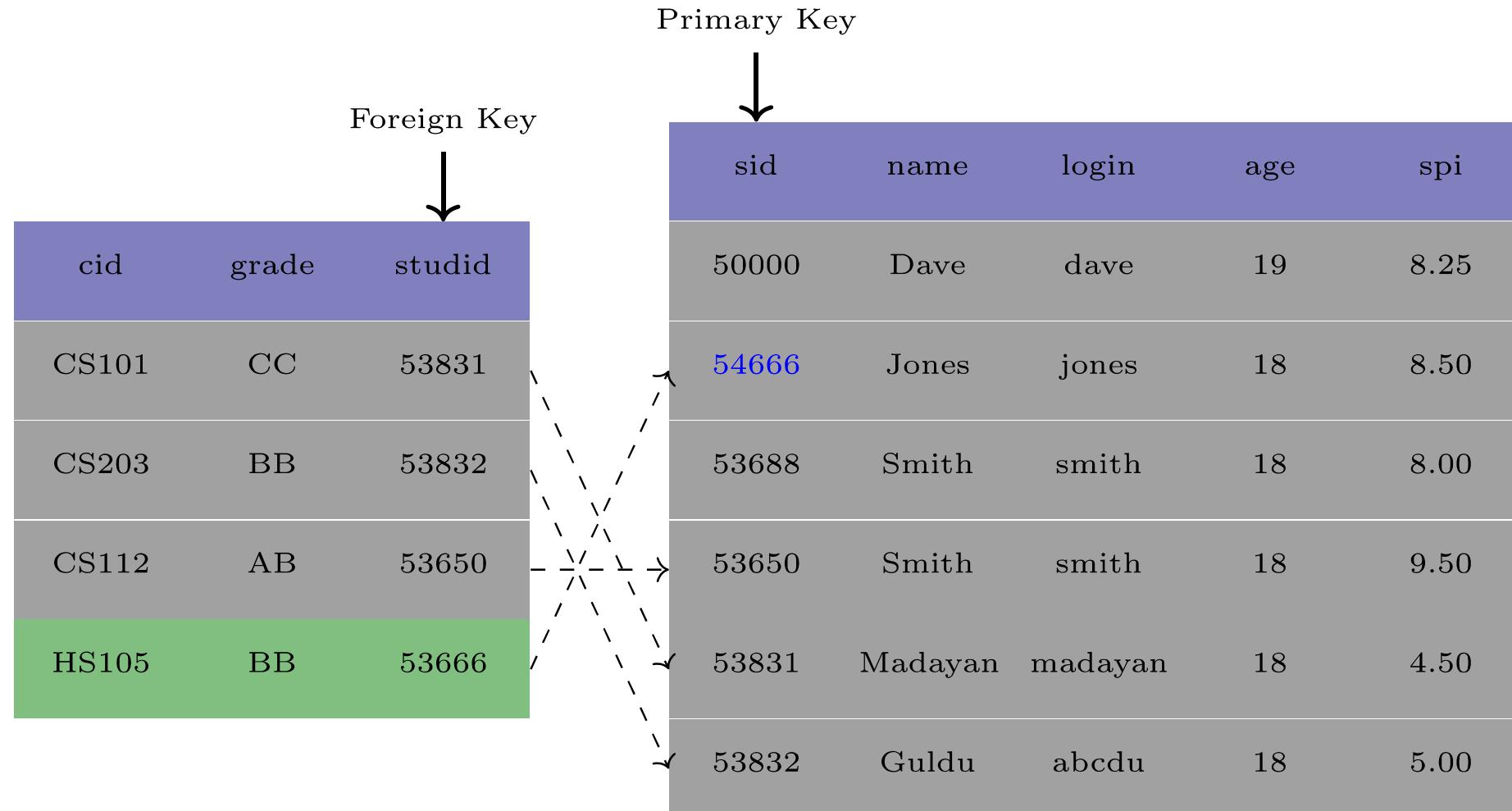
Write a NULL values in all rows in LHS table with studid=53666

Foreign Key			Primary Key				
cid	grade	studid	sid	name	login	age	spi
CS101	CC	53831	50000	Dave	dave	19	8.25
CS203	BB	53832	53666	Jones	jones	18	8.50
CS112	AB	53650	53688	Smith	smith	18	8.00
HS105	BB	⊥	53650	Smith	smith	18	9.50
			53831	Madayan	madayan	18	4.50
			53832	Guldu	abcd	18	5.00

Foreign Key - Row Deletion from RHS Table - Action 04

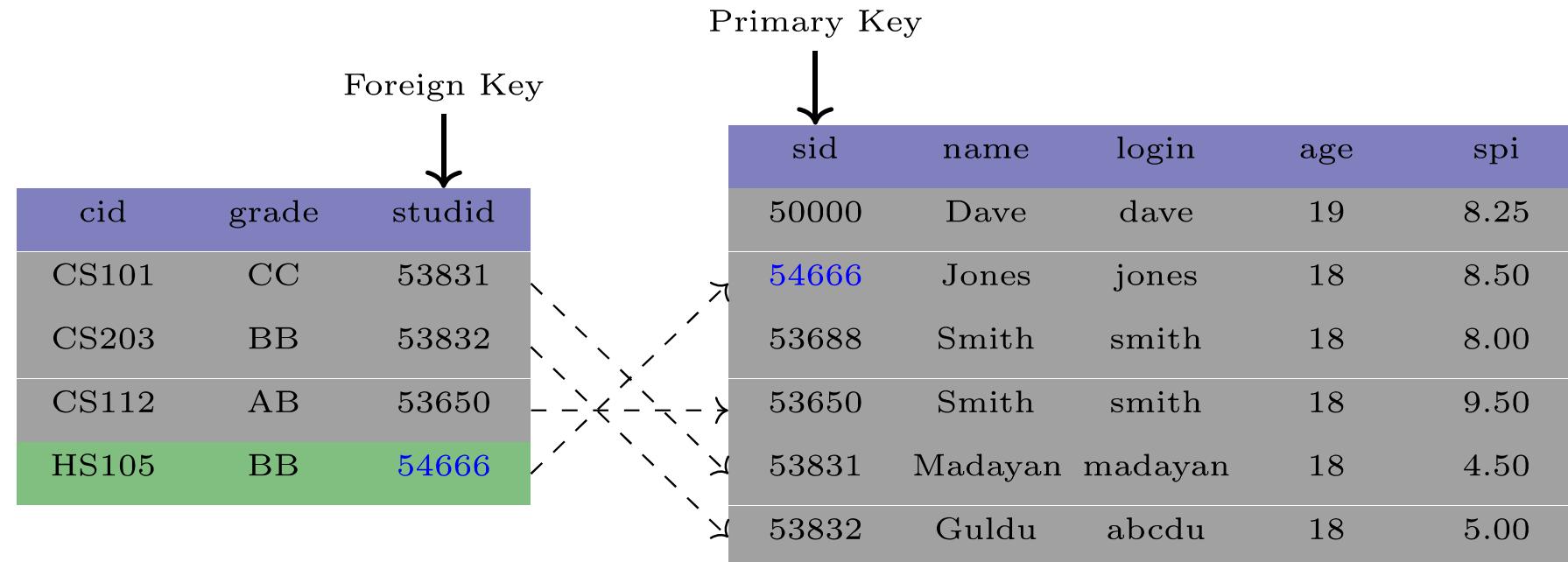


Foreign Key - Updated a row in RHS Table



Foreign Key - Row in RHS Table Updated - Action 01

Update all rows in LHS table with studid=54666



Foreign Key - Row in RHS Table Updated - Action 02

Write a DEFAULT values (say 50000) in all rows in LHS table with studid=53666

Foreign Key			Primary Key				
cid	grade	studid	sid	name	login	age	spi
CS101	CC	53831	50000	Dave	dave	19	8.25
CS203	BB	53832	54666	Jones	jones	18	8.50
CS112	AB	53650	53688	Smith	smith	18	8.00
HS105	BB	50000	53650	Smith	smith	18	9.50
			53831	Madayan	madayan	18	4.50
			53832	Guldu	abcd	18	5.00
			99999	Dummy	dummy	dummy	dummy

Foreign Key - Row in RHS Table Updated - Action 03

Write a NULL values in all rows in LHS table with studid=53666

Foreign Key			Primary Key			
cid	grade	studid	sid	name	login	age
CS101	CC	53831	50000	Dave	dave	19
CS203	BB	53832	54666	Jones	jones	18
CS112	AB	53650	53688	Smith	smith	18
HS105	BB	⊥	53650	Smith	smith	18
			53831	Madayan	madayan	18
			53832	Guldu	abcd	5.00

Foreign Key - Row in RHS Table Updated - Action 04

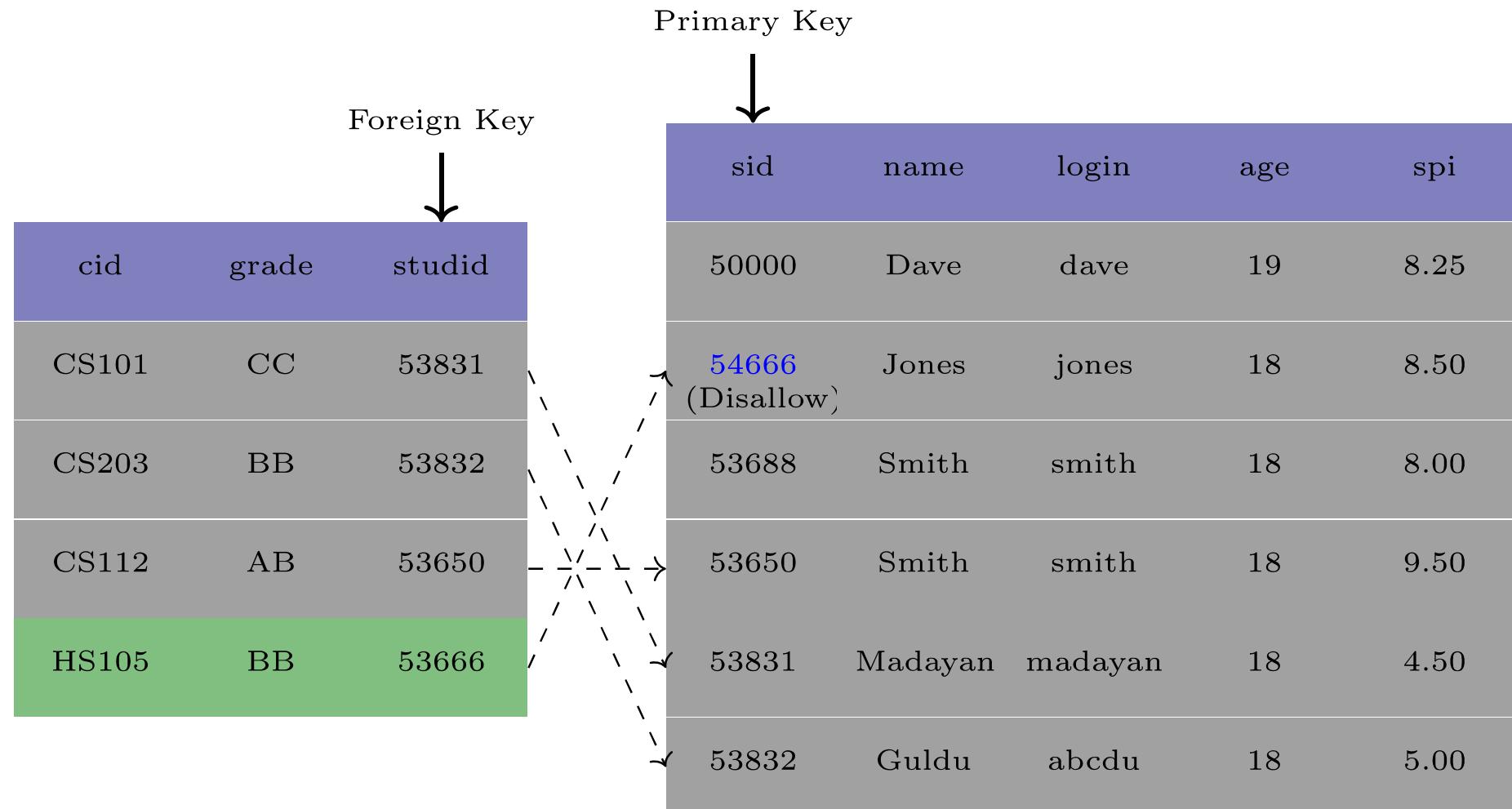


Table operations

Single Table

- Discussing operations on a single table
- Create table with specified number of columns, their names and their data types
- Delete table
- Add a column to the existing table (at the beginning)
- Add a column to the existing table (in the middle)
- Add a column to the existing table (at the end)
- Delete a column from an existing table
- Change column data type

Single Table

- Add a constraint to existing table
- Delete an existing constraint from a table

Brief history

Brief History

- E. F. Codd (IBM Research Laboratory) invented the [Relational Databases](#)
- Was awarded Turing award in 1981 for the seminal work
- Try to read the paper [A Relational Model of Data for Large Shared Data Banks](#)
- A theoretical model defining relations, and operations on relations
- Followed by 12 rules of Codd for the relational databases

Brief History

- 1974: IBM's [System R](#) prototype of RDBMS
- 1979: Oracle Corporation's [Oracle](#)
- 1970: [Ingres](#) by University of California
- 1987: SAP by Sybase
- Try to read the paper [A Relational Model of Data for Large Shared Data Banks](#)
- A theoretical model defining relations, and operations on relations
- Followed by 12 rules of Codd for the relational databases

Brief History

- Structured Query Language (SQL)
- Designed for managing data in RDBMS
- first version is known as SEQUEL (Structured English Query Language)
- Developed by [Donald D. Chamberlin](#) and [Raymond F. Boyce](#)
- First standard is available in the year 1986 formalized by ANSI (SQL-86)
- Latest standard is published in 2019 (SQL:2019)

SQL

Overview

DDL Subset of SQL support **creation, deletion** and **modification** of tables and views

DML Subset of SQL that allows users to pose **queries, insert, delete and modify** tuples

Triggers, Events & Adv. Constraints Performs operations based on actions or time

Embedded SQL SQL statements can be included in various programming languages such as C, C++, Java, python and/or php

Overview

Transaction Management Various commands allow user to explicitly control aspects of how a transaction is to be executed

Security provide mechanism to control user's access to tables and views

Programming Constructs such as control statements, loops, exceptions, error handling statements