DALHOUSIE UNIVERSITY

Data Management, And Warehousing-Analytics

Lab 2

Summer 2024

DATA MODELLING PROCESS

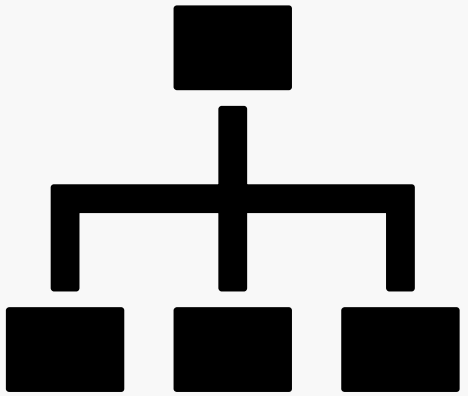
Teaching Assistants & Markers
Axata Darji(ax583820@dal.ca)
Bharat Shankaranarayanan (bh277785@dal.ca)
Mahsa Mousavi Diva(mahsa.msv@dal.ca)

Some content in this presentation is referred from Winter 2024 (with permission)

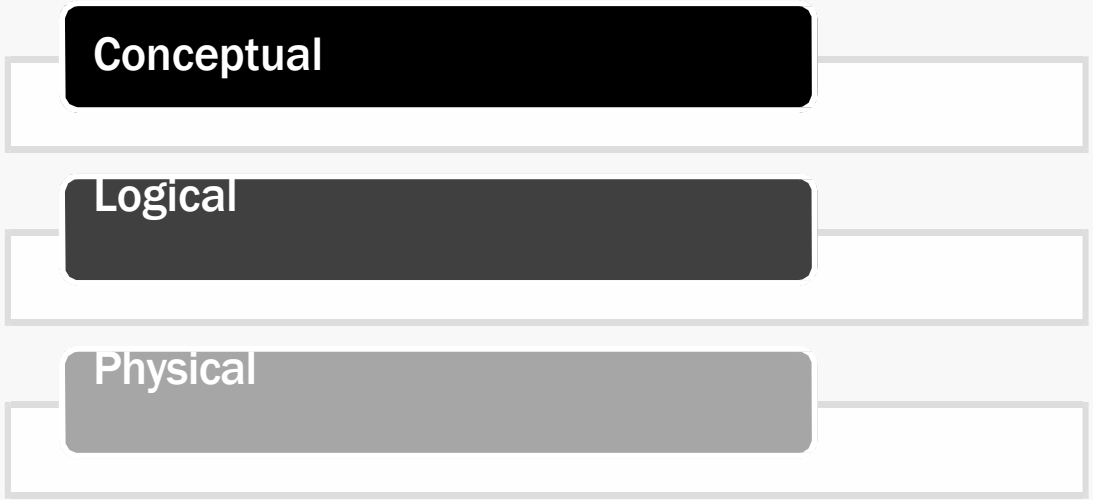
Data modelling

Data modelling

- *A process for defining **fundamental components** of the database*
- **It is the process of creating a data model for the data to be stored in a database**



Phases of Data Modelling



Data modelling flowchart

Gather	Gather enough requirements
Identify	Identify entities and attributes
Design	Design a basic conceptual model
Design	Design a logical model
Create	Finally, create the database/ Physical model

Normalization

- Process of organizing the database.
- Main objective of normalization is to:
  1. *Reduce the redundancy,*
  2. *Simplify the queries and reduce the cost,*
  3. *Reduce/Avoid data modification issues.*

Normalization Forms

Normal Form	Description
First Normal Form (1NF)	Should be in a proper format, no group values for a field and a primary key is identified.
Second Normal Form (2NF)	Should be in 1st Normal Form and there should be no partial dependency.
Third Normal Form (3NF)	Should be in 2nd Normal Form and there should be no transitive dependency.

1 NF Example

FULL NAMES	PHYSICAL ADDRESS	MOVIES RENTED	SALUTATION
Janet Jones	First Street Plot No 4	Pirates of the Caribbean, Clash of the Titans	Ms.
Robert Phil	3rd Street 34	Forgetting Sarah Marshal, Daddy's Little Girls	Mr.
Robert Phil	5th Avenue	Clash of the Titans	Mr.

Source: What is Normalization in DBMS (SQL)? 1NF, 2NF, 3NF Example (guru99.com)

1 NF Example

FULL NAMES	PHYSICAL ADDRESS	MOVIES RENTED	SALUTATION
Janet Jones	First Street Plot No 4	Pirates of the Caribbean	Ms.
Janet Jones	First Street Plot No 4	Clash of the Titans	Ms.
Robert Phil	3rd Street 34	Forgetting Sarah Marshal	Mr.
Robert Phil	3rd Street 34	Daddy's Little Girls	Mr.
Robert Phil	5th Avenue	Clash of the Titans	Mr.

Source: What is Normalization in DBMS (SQL)? 1NF, 2NF, 3NF Example (guru99.com)

Partial Dependency

- When we have a composite primary key and some attributes dependent only on a part of the primary key then it is a partial dependency.

2 NF Example

FULL NAMES	PHYSICAL ADDRESS	MOVIES RENTED	SALUTATION
Janet Jones	First Street Plot No 4	Pirates of the Caribbean	Ms.
Janet Jones	First Street Plot No 4	Clash of the Titans	Ms.
Robert Phil	3rd Street 34	Forgetting Sarah Marshal	Mr.
Robert Phil	3rd Street 34	Daddy's Little Girls	Mr.
Robert Phil	5th Avenue	Clash of the Titans	Mr.

Source: What is Normalization in DBMS (SQL)? 1NF, 2NF, 3NF Example (guru99.com)

2 NF Example

MEMBERSHIP ID	FULL NAMES	PHYSICAL ADDRESS	SALUTATION
1	Janet Jones	First Street Plot No 4	Ms.
2	Robert Phil	3rd Street 34	Mr.
3	Robert Phil	5th Avenue	Mr.

MEMBERSHIP ID	MOVIES RENTED
1	Pirates of the Caribbean
1	Clash of the Titans
2	Forgetting Sarah Marshal
2	Daddy's Little Girls
3	Clash of the Titans

Source: What is Normalization in DBMS (SQL)? 1NF, 2NF, 3NF Example (guru99.com)

Transitive Dependency

- When an attribute is independent of the primary key but dependent on some other non-key attribute then it is transitive dependency.

3 NF Example

MEMBERSHIP ID	FULL NAMES	PHYSICAL ADDRESS	SALUTATION
1	Janet Jones	First Street Plot No 4	Ms.
2	Robert Phil	3rd Street 34	Mr.
3	Robert Phil	5th Avenue	Mr.

MEMBERSHIP ID	MOVIES RENTED
1	Pirates of the Caribbean
1	Clash of the Titans
2	Forgetting Sarah Marshal
2	Daddy's Little Girls
3	Clash of the Titans

Source: What is Normalization in DBMS (SQL)? 1NF, 2NF, 3NF Example (guru99.com)

3 NF Example

MEMBERSHIP ID	FULL NAMES	PHYSICAL ADDRESS	SALUTATION ID
1	Janet Jones	First Street Plot No 4	2
2	Robert Phil	3rd Street 34	1
3	Robert Phil	5th Avenue	1

MEMBERSHIP ID	MOVIES RENTED
1	Pirates of the Caribbean
1	Clash of the Titans
2	Forgetting Sarah Marshal
2	Daddy's Little Girls
3	Clash of the Titans

SALUTATION ID	SALUTATION
1	Mr.
2	Ms.
3	Mrs.
4	Dr.

Source: What is Normalization in DBMS (SQL)? 1NF, 2NF, 3NF Example (guru99.com)

De-Normalization

- Refers to the process of intentionally introducing redundancy into a table structure to improve query performance
- The purpose of denormalization is to reduce the number of Joins as Joins can be expensive and slow, especially when dealing with large and complex databases.
- By de-normalizing the database, you can store frequently accessed or related data in one table, which can improve the query speed and response time.

De-Normalization Challenges

- It can increase storage space requirements, as redundant data or combined tables will need to be stored
- Update anomalies may arise when updating the data in a denormalized database, leading to insertion, deletion, or modification issues that affect the integrity and accuracy of the data.

When to do De-Normalization?

- Depends on the use case and the trade-offs involved.
- Frequent read operations and infrequent write operations
- Query efficiency rather than transactional consistency.

Transactions

- A Database transaction is a feature in the database which helps to perform one or more operations maintaining the consistency of the database.
- All the statements between the beginning and ending of a transaction can be considered as a single unit.
- While performing a Transaction, the database will be in an inconsistent state.
- Only if the transaction is committed (COMMIT;) the database's state is changed from one consistent state to other.
- Transactions that do not modify data in the tables but only fetch the required data are known as read-only transactions.
- ACID properties.

ACID properties in DBMS

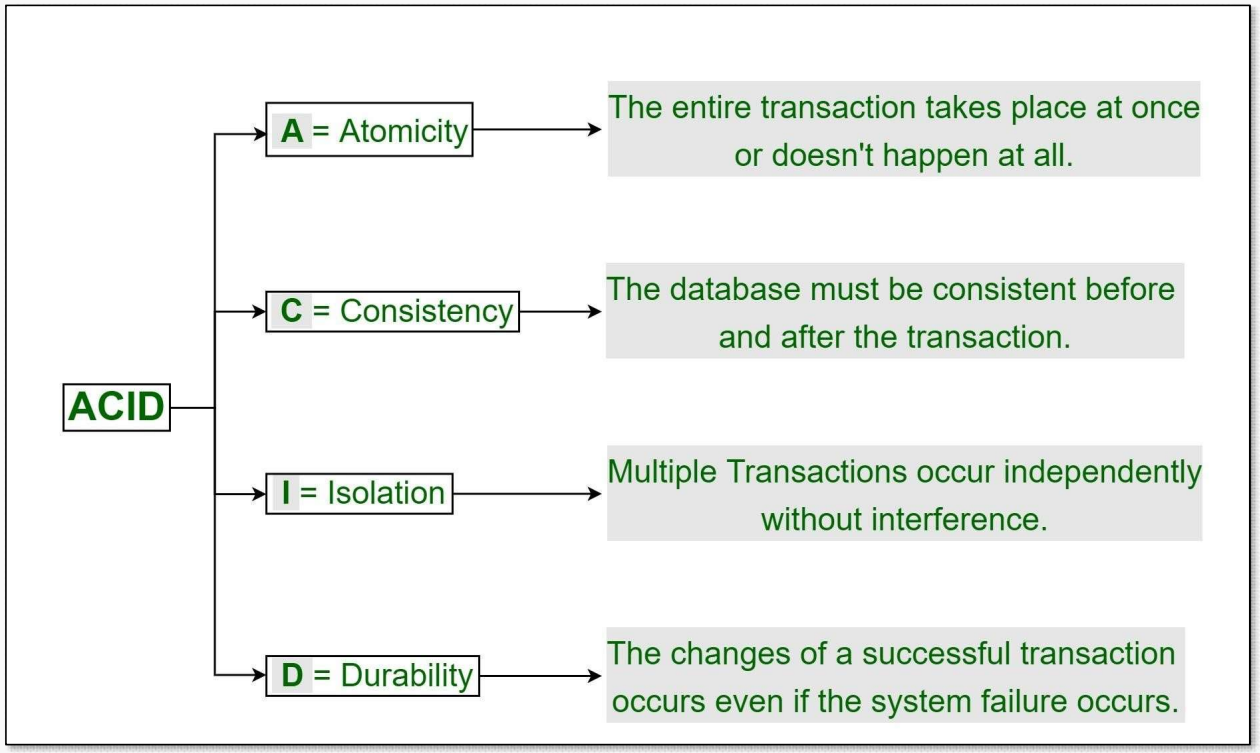


Image source: https://www.geeksforgeeks.org/acid-properties-in-dbms/

Transaction Syntax

- To inform the system that a transaction is being started, we use START TRANSACTION command.
- We can perform any operation which we need inside a transaction UPDATE, INSERT,SELECT, etc.
- At the end of the transaction, we can either COMMIT the transaction or ROLLBACK the transaction.

Transaction Syntax continued…

- SET autocommit = 0;
- before the start of the transaction.
- Example:

```
START TRANSACTION;

    INSERT INTO Student VALUES(1, "XYZ", "XYZ@dal.ca");
    INSERT INTO Student VALUES(2, "ABC", "ABC@dal.ca");
    INSERT INTO Student VALUES(3, "PQR", "PQR@dal.ca");

COMMIT; ROLLBACK;
```

Transaction Lifecycle: Different states in a Transaction

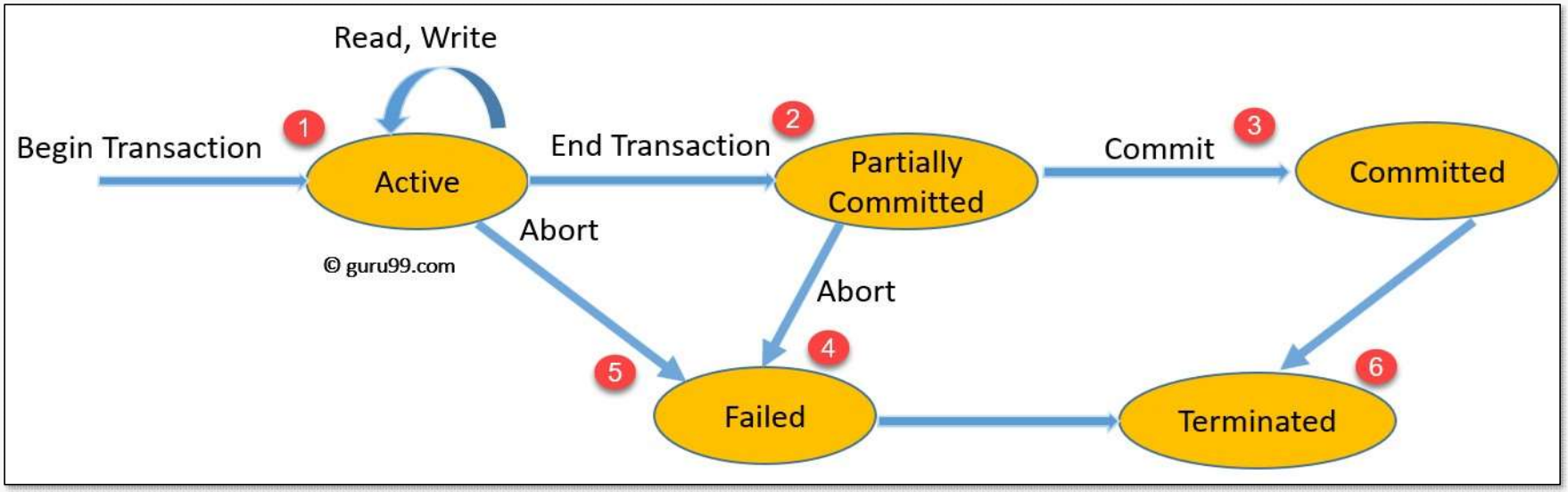
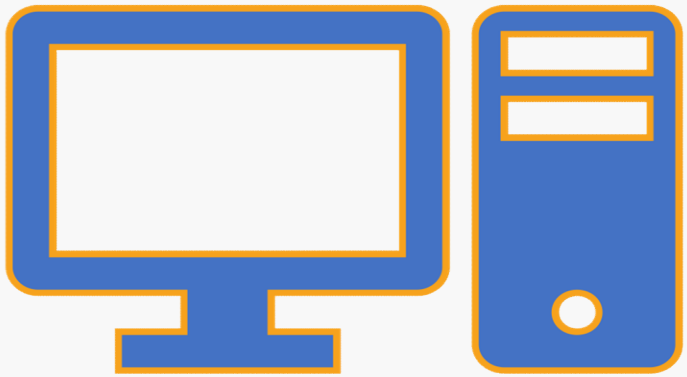


Image source : https://www.guru99.com/dbms-transaction-management.html

SAVEPOINT

- SAVEPOINT is like a flag which breaks down a task into several tasks.
- SAVEPOINT helps to rollback a part of the transaction.
- SAVEPOINTs are exactly like the checkpoints in the video games. We create a checkpoint when we have completed a sure or difficult task.
- The SAVEPOINTs are released after a COMMIT or ROLLBACK. However, we can use RELEASE SAVEPOINT to manually release a SAVEPOINT.
- In case of RELEASE, the SAVEPOINT is destroyed without undoing the effects of the queries executed after the SAVEPOINT was created. You cannot ROLLBACK a SAVEPOINT after RELEASE.
- In case of ROLLBACK, all queries after the SAVEPOINT are rolled back without destroying the rest of the transaction

Time for some hands-on (Transactions)



Example using SAVEPOINT

```
4
5  INSERT INTO employee(emp_id, emp_name)
6  VALUES (6, 'Rita');
7
8  SAVEPOINT my_savepoint;
9
10 INSERT INTO employee(emp_id, emp_name)
11 VALUES (7, 'Jeremy');
12
13 ROLLBACK TO SAVEPOINT my_savepoint;
14
15 INSERT INTO employee(emp_id, emp_name)
16 VALUES (8, 'Jonathan');
17
18 SELECT * FROM employee;
19
20 COMMIT;
```

Transaction with rolled back Savepoint

Example using SAVEPOINT

```
1  START TRANSACTION;
2  INSERT INTO employee(emp_id, emp_name)
3  VALUES (9, 'Jimmy');
4
5  SAVEPOINT my_savepoint;
6
7  update employee set emp_name = 'Saville' where emp_id=9;
8
9  RELEASE SAVEPOINT my_savepoint;
10
11 INSERT INTO employee(emp_id, emp_name)
12 VALUES (11, 'Jake');
13
14 SELECT * FROM employee;
15
16 COMMIT;
```

Transaction with released Savepoint

Graded Exercise

Question: Normalize Invoice Table

InvoiceID	Date	Branch	BranchLocation	City	CustomerType	CustomerName	Gender	Product	UnitPrice	Quantity
1101	2/5/2024	1,004,008	Quinpool,Windsor,Dartmouth	Halifax	Member,Normal	Alex,Rohan,Mark	Female,Male	Phone,Battery,Cover	200,50,25	1,2,5
2104	3/3/2024	3,007,008	Dowry Street, Abbey Lane, GlenView	Toronto	Member,Normal	Suzan,Carla,Andreq	Female,Male	Code,Power bank	10,35	20,45

1. *Normalize Invoice table to **1NF**.*
2. *Normalize the 1NF tables to **2NF.***
3. *Normalize 2NF tables to **3NF** .*

At each normal form stage, present the resulting tables with their columns and dependencies.

4. Do you think it will be a good design decision to de-normalize the normal forms generated for the Product table in the above steps? Why/Why not? Provide a brief explanation.

Deliverables

- ***For Steps 1,2 and 3***
  - *Provide diagrams showing the dependencies in the table(where applicable). This can be a set of arrows indicating the dependencies between columns(see Figure 1).*

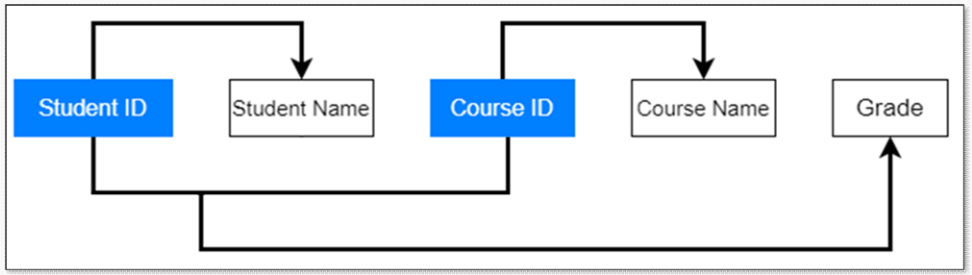


Figure 1

  - *Give a brief explanation of the normalization process at each step, discussing why specific dependencies were identified(where applicable) and how the table was decomposed to adhere to that normal form.*

- ***For Step 4***
  - *Provide a brief explanation in paragraph or bullet point form.*

Deadline: Before Sunday(19th May 2024), 11:59 pm. Please refer "Submission Guideline" in Brightspace.

Q&A

References

- D. Taylor, “What is data modelling? types (conceptual, logical, physical),” *Guru99*, 05-Jan-2023. [Online]. Available: https://www.guru99.com/data-modelling-conceptual-logical.html. [Accessed: 14-Sep-2023].
- “Database model,” *Wikipedia*, 23-Nov-2022. [Online]. Available: https://en.wikipedia.org/wiki/Database_model. [Accessed: 14-Sep-2023].
- “What is the purpose of denormalization in relational databases?”, LinkedIn[Online]. Available: https://www.linkedin.com/advice/0/what-purpose-denormalization-relational-databases-2acff. [Accessed: January 23,2024].
- “DBMS Transaction Management: ACID Properties, Schedule,” Guru99 [Online]. Available at: https://www.guru99.com/dbms-transaction-management.html [Accessed 24 September 2023].
- “ACID Properties in DBMS - GeeksforGeeks,” GeeksforGeeks [Online]. Available at: https://www.geeksforgeeks.org/acid-properties-in-dbms/ [Accessed 24 September 2023].