Database Concepts

Wipro Training Feb 2011

TEAM ABAP

Introduction to Database

Evolution of Database

- What is File system?
  - Method of storing and organizing computer files and their data.
  - Way in which files are named and placed logically for storage and retrieval
- Disadvantages:
  - Data Redundancy.
  - Poor data control.
  - Data inconsistency.
  - Security issue.
  - Difficulty in accessing data.
  - Integrity issues.

Database

- What is a Database?

  - collection of organized data on computer
 - Organized mechanism for storing, managing and retrieving information

- What is DBMS?

  - Set of computer programs that controls the creation, maintenance, and the use of a database
  - Software system that uses a standard method of cataloging, retrieving, and running queries on data.

Database Architecture

Architecture of Database

▪**External** : this level defines, how the user understands the organization of data within the database. A single database can have any number of views in the external level.

▪**Internal**: this level defines how the data is physically stored and how the system computes the data.

▪**Conceptual** : It unifies the various external views to make the database more meaningful.
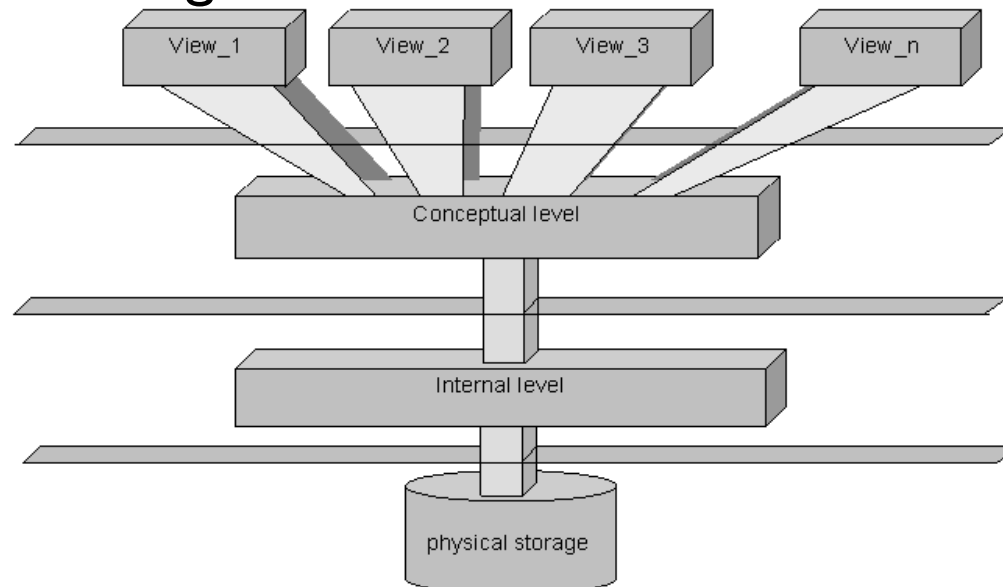
The Three ANSI/SPARC Levels of Abstractions

Database Schema

Database Schema

- A Database Schema is described as a formal language supported by the Database Management System (DBMS).

- In a relation database, the schema defines the table, fields, and the relationship between the fields of the different tables.

- **Conceptual schema** : Refers to the entities and attributes that are to be considered for an organization and the relationship between them, shows how real world entities are modeled in db

- **Logical schema** : Deals with the representation of data, whether it conforms to the constraints of particular approach of DBMS.

- **Physical schema** : Deals with the implementation of logical schema and physical storage of data.

Benefits of Database Schema

- Logical Data Independence
  - The capacity to change the conceptual schema without having to change the external schema or application program.

- Physical data Independence
  - The capacity to change the internal schema without having to change the conceptual or external schema.
  - Internal schema may change to improve the performance.

Database Model

Database Model

- It is collection of logical structure used to represent data structure and relationship within database.

- Types of Database Model

  - Hierarchical Database Model:

    - Tree like Structure.

    - One to Many relationship.

  - Network Database Model:

    - Many to many relationship.

    - Tree like Structure

Database Model

- Relational Database Model:

    - Uses a collection of tables to represent both data and relationship among those data.

- Object-oriented Database Model:

    - The data is stored in the form of objects, which are structures called classes.

Database Query

DATABASE QUERY

- A database query is a piece of code (a query)that is used for retriving the information from database.

- A database "query" is basically a "question" that you ask the database. The results of the query is the information that is returned by the database management system

- Queries are usually constructed using SQL (structured query language) which resembles a high-level programming language.

DATABASE STATEMENTS

- DDL(Date Definition Language)
  - Used to build and modify the structure of tables and other objects in the database.

- DML(DATA MANIPULATIONLANGUAGE)
  - Used to store, retrieve, modify, and erase data from a database.

Data definition language consist of

- CREATE - to create objects in the database
- ALTER - alters the structure of the database
- DROP - delete objects from the database
- TRUNCATE - remove all records from a table, including all spaces allocated for the records are removed

Data Manupulation language consist of

- SELECT - retrieve data from the a database
- INSERT - insert data into a table
- UPDATE - updates existing data within a table
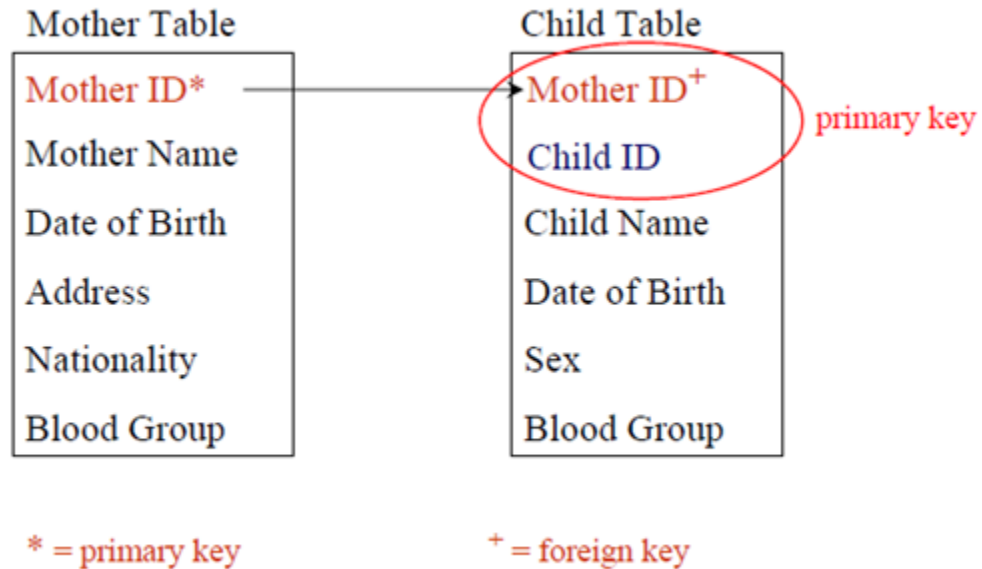- DELETE - deletes all records from a table, the space for the records remain .

Key Concepts

Types of Keys

- Ensure each record within a table can be identified by one or a combination of fields within the table.

  - Super Key
  - Candidate Key
  - Primary Key
  - Foreign Key

Mother Table	Child Table
Mother ID*	Mother ID+
Mother Name	Child ID
Date of Birth	Child Name
Address	Date of Birth
Nationality	Sex
Blood Group	Blood Group

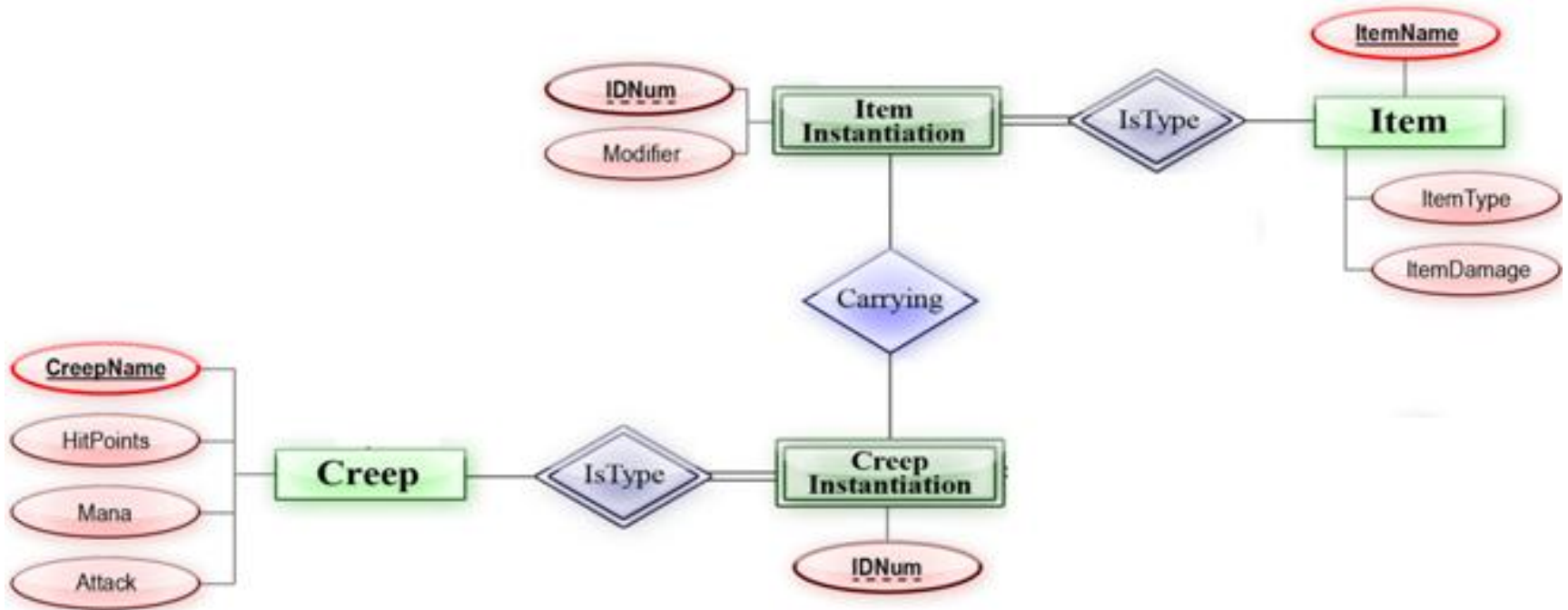primary key

* = primary key   + = foreign key

Database Concepts

Entity Relationship

- It is based on entities, attributes, and relationships.

- A database can be modeled as: a collection of entities, relationship among entities.

- An **entity** is an object that exists and is distinguishable from other objects.

- Entities have attributes

Entity model

Weak Entity

- An entity set that does not have a primary key is referred to as a weak entity set.

- The discriminator (or partial key) of a weak entity set is the set of attributes that distinguishes among all the entities of a weak entity set.

- The primary key of a weak entity set is formed by the primary key of the strong entity set on which the weak entity set is existence dependent, plus the weak entity set's discriminator.
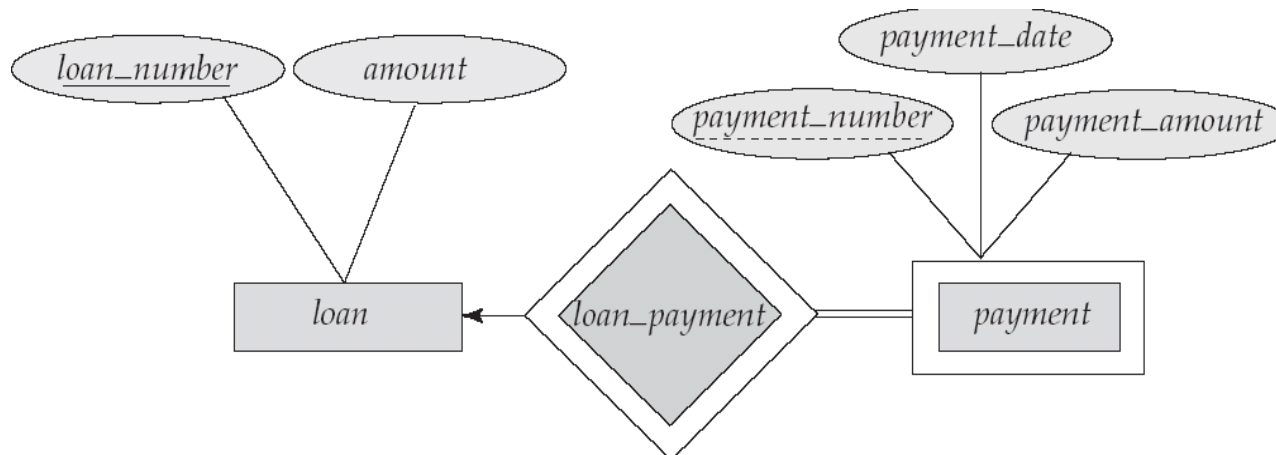
Views And Joins

- Used for selecting fields that are scattered across different tables.

- Tables from which fields are selected, referred to as Base tables or Underlying tables.

- They don't occupy memory.

- Views are implemented by Joins.

Inner Join

- The INNER JOIN return rows when there is at least one match in both tables

Employee

EmployeeID	Name	Telephone	StartedEmployment
1	Bob Marley	0222 00000	10/01/2005
2	John Lennon	0222 00050	02/05/2003
3	Ralph Kimball	0222 03307	01/04/2004
4	Bill Gates	0222 03307	01/04/2004

TrainingTaken

TrainingTakenID	EmployeeID	TrainingTitle	TrainingDate
1	2	Sales Training	10/01/07
2	1	Risk Management	05/02/07
3	2	First Aid	01/03/07
4	3	Sales Training	10/01/07

Name	TrainingTaken	TrainingDate
Bob Marley	Risk Management	05/02/07
John Lennon	Sales Training	10/01/07
John Lennon	First Aid	01/03/07
Ralph Kimball	Sales Training	10/01/07

SELECT Employee.Name,TrainingTaken.TrainingTitle,TrainingTaken.TrainingDate FROM Employee
INNER JOIN TrainingTaken ON Employee.EmployeeID = TrainingTaken.EmployeeID

Outer join

- OUTER JOIN allows us to retrieve all values in a certain table regardless of whether these values are present in other tables.

REGNO	MAKE	COLOUR	PRICE	OWNER
F611 AAA	FORD	RED	12000	Jim Smith
J111 BBB	SKODA	BLUE	11000	Jim Smith
A155 BDE	MERCEDES	BLUE	22000	Bob Smith
K555 GHT	FIAT	GREEN	6000	Bob Jones
SC04 BFE	SMART	BLUE	13000	

NAME	DOB
Jim Smith	11 Jan 1980
Bob Smith	23 Mar 1981
Bob Jones	3 Dec 1986

- Select * from car left join driver on (owner = name);
  Or

- Select * from driver right join car on (owner = name);

REGNO	MAKE	COLOUR	PRICE	OWNER	NAME	DOB
F611 AAA	FORD	RED	12000	Jim Smith	Jim Smith	11 Jan 1980
J111 BBB	SKODA	BLUE	11000	Jim Smith	Jim Smith	11 Jan 1980
A155 BDE	MERCEDES	BLUE	22000	Bob Smith	Bob Smith	23 Mar 1981
K555 GHT	FIAT	GREEN	6000	Bob Jones	Bob Jones	3 Dec 1986
SC04 BFE	SMART	BLUE	13000			

Full Outer join

- Adding another record to the driver table.

NAME	DOB
Jim Smith	11 Jan 1980
Bob Smith	23 Mar 1981
Bob Jones	3 Dec 1986
David Davis	1 Oct 1975

REGNO	MAKE	COLOUR	PRICE	OWNER
F611 AAA	FORD	RED	12000	Jim Smith
J111 BBB	SKODA	BLUE	11000	Jim Smith
A155 BDE	MERCEDES	BLUE	22000	Bob Smith
K555 GHT	FIAT	GREEN	6000	Bob Jones
SC04 BFE	SMART	BLUE	13000	

```
SELECT *
FROM    car FULL JOIN driver ON (owner = name)
;
```

REGNO	MAKE	COLOUR	PRICE	OWNER	NAME	DOB
F611 AAA	FORD	RED	12000	Jim Smith	Jim Smith	11 Jan 1980
J111 BBB	SKODA	BLUE	11000	Jim Smith	Jim Smith	11 Jan 1980
A155 BDE	MERCEDES	BLUE	22000	Bob Smith	Bob Smith	23 Mar 1981
K555 GHT	FIAT	GREEN	6000	Bob Jones	Bob Jones	3 Dec 1986
SC04 BFE	SMART	BLUE	13000			
					David Davis	1 Oct 1975

Database Normalization

Database Normalization

- **Database normalization** is the process of removing redundant data from your tables in to improve storage efficiency, data integrity, and scalability.

- Normalization generally involves splitting existing tables into multiple ones, which must be re-joined or linked each time a query is issued.

- Edgar F. Codd first proposed the process of normalization and what came to be known as the **1st normal form** in his paper *A Relational Model of Data for Large Shared Data Banks* Codd

Benefits of Database Normalization

- Decreased storage requirements!

	1 VARCHAR(20)
converted to	1 TINYINT UNSIGNED
in a table of	1 million rows
is a savings of	~20 MB

- Faster search performance!
    - Smaller file for table scans.
    - More directed searching.
- Improved data integrity!

Functional Dependencies

- What is Functional Dependency:

  - Functional Dependencies are fundamentals to the process of Normalization.

  - It concerns the dependence of the values of one set of attributes on those of another set of attributes.

- Example: **BookId → BookName**

- Types of Functional Dependency

  - Fully Functional Dependency.

  - Partial Functional Dependency.

  - Transitive Functional Dependency.

Table 1

Title	Author1	Author 2	ISBN	Subject	Pages	Publisher
Database System Concepts	Abraham Silberschatz	Henry F. Korth	0072958863	MySQL, Computers	1168	McGraw-Hill
Operating System Concepts	Abraham Silberschatz	Henry F. Korth	0471694665	Computers	944	McGraw-Hill

First Normal Form

- Remove horizontal redundancies
  - No two columns hold the same information
  - No single column holds more than a single item
- Each row must be unique
  - Use a primary key
- Benefits
  - Easier to query/sort the data
  - More scalable
  - Each row can be identified for updating

First Normal Form

- Table 2

Title	Author	ISBN	Subject	Pages	Publisher
Database System Concepts	Abraham Silberschatz	0072958863	MySQL	1168	McGraw-Hill
Database System Concepts	Henry F. Korth	0072958863	Computers	1168	McGraw-Hill
Operating System Concepts	Henry F. Korth	0471694665	Computers	944	McGraw-Hill
Operating System Concepts	Abraham Silberschatz	0471694665	Computers	944	McGraw-Hill

Second Normal Form

- Table must be in First Normal Form
- Remove vertical redundancy
  - The same value should not repeat across rows
- Composite keys
  - All columns in a row must refer to BOTH parts of the key
- Benefits
  - Increased storage efficiency
  - Less data repetition

Subject Table

Subject_ID	Subject
1	MySQL
2	Computers

Author Table

Author_ID	Last Name	First Name
1	Silberschatz	Abraham
2	Korth	Henry

Book Table

ISBN	Title	Pages	Publisher
0072958863	Database System Concepts	1168	McGraw-Hill
0471694665	Operating System Concepts	944	McGraw-Hill

Relationships

Book_Author Table

ISBN	Author_ID
0072958863	1
0072958863	2
0471694665	1
0471694665	2

Book_Subject Table

ISBN	Subject_ID
0072958863	1
0072958863	2
0471694665	2

Third Normal Form

- Table must be in Second Normal Form
    - If your table is 2NF, there is a good chance it is 3NF
- All columns must relate directly to the primary key

- Benefits
    - No extraneous data

Team ABAP Feb 2011