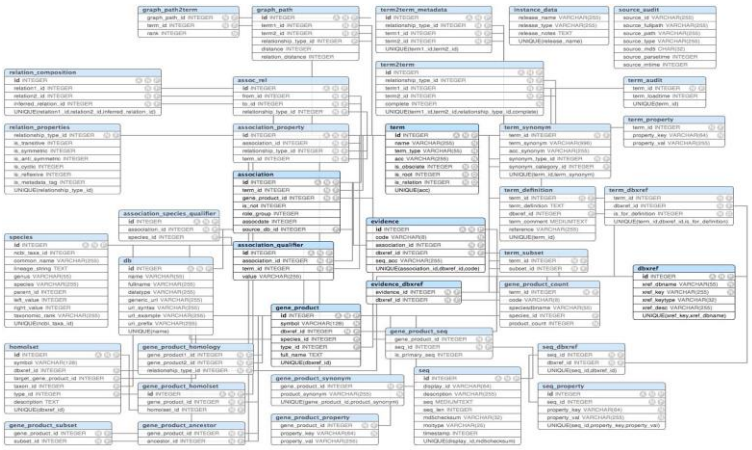
Database chapter 1



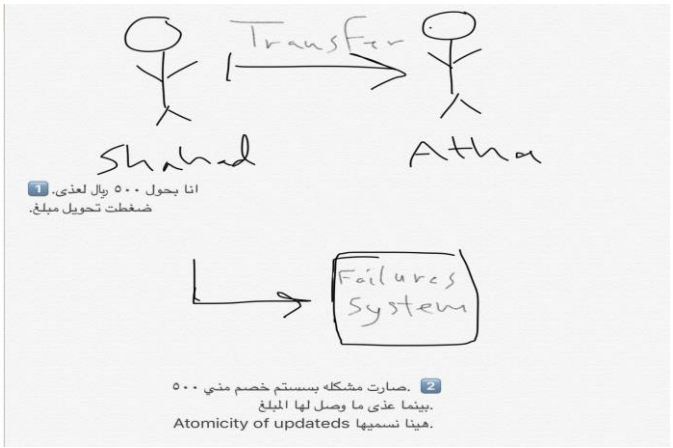
- **Database Management System (DBMS)**
- DBMS contains information about a particular enterprise
  - Collection of Interrelated data ' مجموعه من البيانات المترابطة '
  - Set of programs to access the data ' مجموعة من البرامج توصل البيانات'
  - An environment that is both convenient and efficient to use ' ملائمة ومكافئه للاستخدام '
- Database Application: ' تطبيقات تستخدم قواعد البيانات'
  1. Banking: transactions 'صفقات'
  2. Airlines: reservation , schedules ' الحجوزات ، الجداول '
  3. Universities: registration , grades 'التسجيل ، الدرجات'
  4. Sales: customers, products, purchases 'العملاء، المنتجات، المشتريات'
  5. Online retailers: order tracking, customized recommendations 'تتبع طلب، توصيات مخصصه'
  6. Manufacturing: production, inventory, orders, supply chain 'الانتاج و المخزون وأوامر و سلسة التوريد'
  7. Human resource: employee records, salaries, tax deductions 'سجلات الموظفين، والرواتب والتخفيضات الضريبية'
- Databases can be very large.
- Databases touch all aspects of our lives
- **University Database Example .** 'يوضح طريقة عمل احد التطبيقات الي تستخدم الداتابيس'
- Application program examples:
  - Add new students, instructors, and courses.
  - Register students for courses, and generate class rosters.
  - Assign grades to students, compute grade point averages (GPA) and generate transcripts
- **Drawbacks of using file systems to store data important :** 'بسابق قبل ماينوجد الداتابيس كان فيه شي اسمه الفايل سستم وهينا يوضح عيوبها بتخزين البيانات و الداتابيس توفرها '
  1. Data redundancy and inconsistency 'تكرار البيانات و تناقض البيانات'

SHAHAD ALHAWASHI

- Multiple file formats, duplication of information in different files 'تنسيقات الملفات المتعددة، تكرار المعلومات في ملفات مختلفة'

2. Difficulty in accessing data 'صعوبة في الوصول الى البيانات'
   - Need to write a new program to carry out each new task 'تحتاج إلى كتابة برنامج جديد لتنفيذ كل مهمة جديدة'
3. Data isolation 'عزل البيانات'
   - multiple files and formats.
4. Integrity problems 'مشاكل السلامة'
   - Integrity constraints (e.g., account balance > 0) become "buried" in program code rather than being stated explicitly. 'قيود التكامل (على سبيل المثال، رصيد الحساب> 0) تصبح "دفن" في كود البرنامج بدلا من صراحة'
   - Hard to add new constraints or change existing ones. 'على إضافة قيود جديدة أو تغيير القائم منها'

5. Atomicity of updates. 'ماحصلت ترجمه توضح المعنى برفق مثال لتوضيحها'
   - Failures may leave database in an inconsistent state with partial updates carried out. '
   - Example: Transfer of funds from one account to another should either complete or not happen at all.



6. Concurrent access by multiple users. 'الوصول المتزامن من قبل عدة مستخدمين'
   - Concurrent access needed for performance. 'لوصول المتزامن اللازمة لأداء'
   - Uncontrolled concurrent accesses can lead to inconsistencies. 'المداخل المتزامنة لا يمكن السيطرة عليها يمكن أن يؤدي إلى تناقضات'
     - Example: Two people reading a balance (say 100) and updating it by withdrawing money (say 50 each) at the same time
7. Security problems

- Hard to provide user access to some, but not all, data. '
- Database systems offer solutions to all the above problems

❖ **Levels of Abstraction: Very important** 'هينا يوضح لك طبقات عمل قواعد البيانات'

1) **Physical level**: describes how a record (e.g., customer) is stored.
2) **Logical level:** describes data stored in database, and the relationships among the data. 'يصف البيانات المخزنة في قاعدة البيانات، والعلاقات بين البيانات'
3) **View level:** application programs hide details of data types. Views can also hide information (such as an employee's salary) for security purposes. 'برامج التطبيقات إخفاء تفاصيل عن أنواع البيانات. آراء يمكن أيضا إخفاء المعلومات (مثل راتب الموظف) لأغراض أمنية.'

❖ Instances and Schemas

Note: Similar to types and variables in programming languages
Schema = is a variable
Instance = is a value of a variable
قواعد البيانات على شكل جدول ، السكيما تمثل الأعمدة
اسم المتغير بينما انستانس تمثل القيامه الي في الصفوف

Schema

Instanc

ID	*name*	*dept_name*	*salary*
22222	Einstein	Physics	95000
12121	Wu	Finance	90000
32343	El Said	History	60000
45565	Katz	Comp. Sci.	75000
98345	Kim	Elec. Eng.	80000
76766	Crick	Biology	72000
10101	Srinivasan	Comp. Sci.	65000
58583	Califieri	History	62000
83821	Brandt	Comp. Sci.	92000
15151	Mozart	Music	40000
33456	Gold	Physics	87000
76543	Singh	Finance	80000

- Similar to types and variables in programming languages.
- **Schema** – the logical structure of the database. 'سكيما - البنية المنطقية من قاعدة البيانات'
  - Example: The database consists of information about a set of customers and accounts and the relationship between them. 'مثال: تتكون قاعدة البيانات من المعلومات حول مجموعة من العملاء والحسابات والعلاقة بينهما'

SHAHAD ALHAWASHI

- Analogous to type information of a variable in a program. ' المشابهة لكتابة المعلومات متغير في برنامج '
- Physical schema: database design at the physical level - مهم تعرفين أنواع السكيما
- Logical schema: database design at the logical level

- **Instance** – the actual content of the database at a particular point in time. - ' انستانس المضمون الفعلي لقاعدة البيانات عند نقطة معينة من الزمن'
  - Analogous to the value of a variable. ' مماثلة لقيمة متغير '
- **Physical Data Independence** – the ability to modify the physical schema without changing the logical schema. ' الاستقلال البيانات الفيزكيل - المقدرة على تعديل الفيزكيل سكيما دون تغيير لوجيكل سكيما'
  - Applications depend on the logical schema. ' تعتمد التطبيقات على لوجيكل سكيما'
  - In general, the interfaces between the various levels and components should be well defined so that changes in some parts do not seriously influence others. ' بشكل عام، واجهات بين مختلف المستويات ومكونات يجب أن تحدد جيدا بحيث تغييرات في بعض أجزاء لا تؤثر على محمل الجد الآخرين. '

❖ **What are the types of binary relationship?**

1. One to one
2. One to many
3. Many to one
4. Many to many

❖ **Data Models** reading 'مجرد مفاهيم بدون تعاريف بندرسها بالشباتر الجايه'

➢ A collection of tools for describing
  - Data
  - Data relationships . 'علاقات البيانات'
  - Data semantics . 'دلائل البيانات'
  - Data constraints. 'قيود البيانات'

➢ Other older models:
  - Network model
  - Hierarchical model

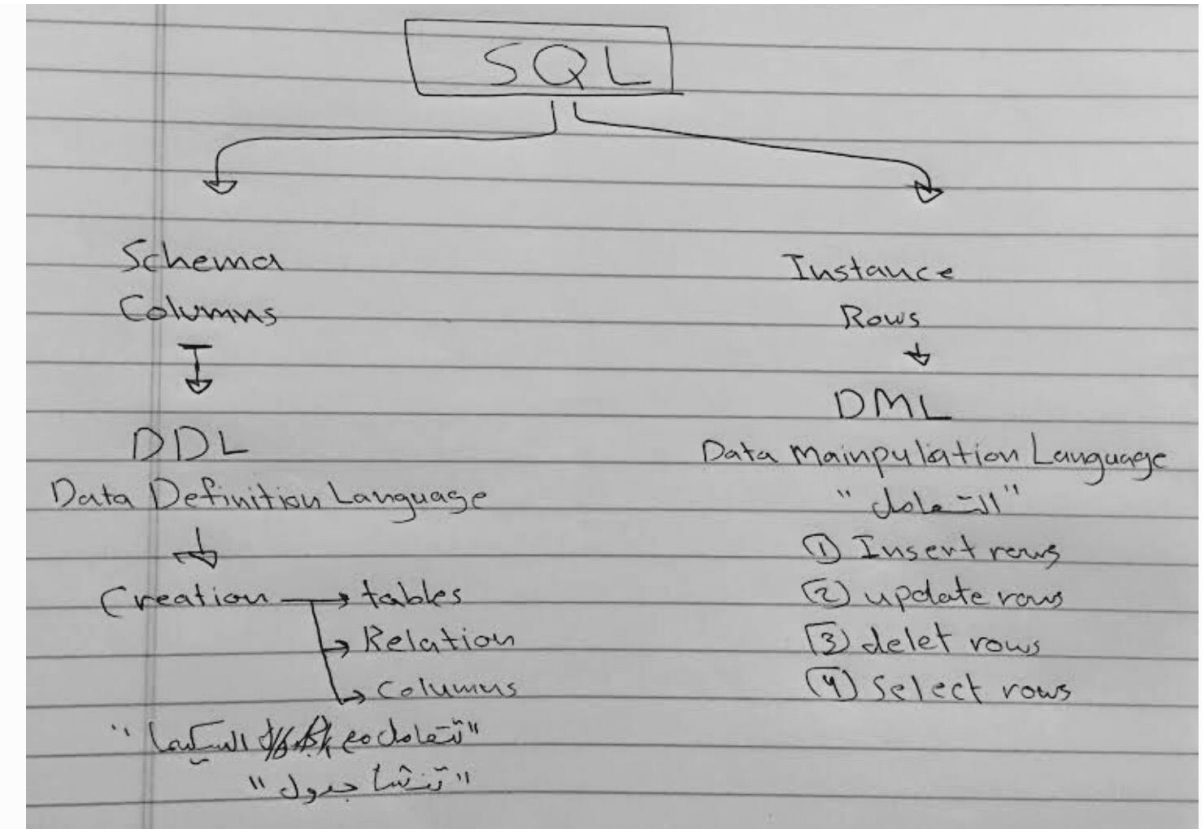
ضروري تعرفين المسميات الي بالهايلايت

Columns

ID	*name*	*dept_name*	*salary*
22222	Einstein	Physics	95000
12121	Wu	Finance	90000
32343	El Said	History	60000
45565	Katz	Comp. Sci.	75000
98345	Kim	Elec. Eng.	80000
76766	Crick	Biology	72000
10101	Srinivasan	Comp. Sci.	65000
58583	Califieri	History	62000
83821	Brandt	Comp. Sci.	92000
15151	Mozart	Music	40000
33456	Gold	Physics	87000
76543	Singh	Finance	80000

Rows

SHAHAD ALHAWASHI



- **Data Manipulation Language (DML)** **Very important** **to know the difference between DML & SQL & DDL**
- Language for accessing and manipulating the data organized by the appropriate data model. ' للغة للوصول الى والتلاعب في البيانات من قبل نموذج البيانات المناسبة
  - DML also known as query language
- What is the difference between procedural and non-procedural languages?
  - **Procedural** – user specifies what data is required and how to get those data. ' يحدد المستخدم ما هو مطلوب البيانات وكيفية الحصول على تلك البيانات'
  - **Declarative (nonprocedural)** – user specifies what data is required without specifying how to get those data. ' يحدد المستخدم ما هو مطلوب البيانات دون تحديد كيفية الحصول على تلك البيانات
- SQL is the most widely used query language . 'هي الأكثر استخداما على نطاق واسع لغة الاستعلام '
- Retrieval of information stored in the database. ' استرجاع المعلومات المخزنة في قاعدة البيانات'
- Insertion of new information into the database. 'الإدراج من معلومات جديدة في قاعدة البيانات '
- Deletion of information from the database ' حذف المعلومات من قاعدة البيانات'
- Modification of information stored in the database. ' تعديل المعلومات المخزنة في الداتابيس'

❖ **Data Definition Language (DDL)**

➢ Specification notation for defining the database schema. ' مواصفات تدوين لتحديد مخطط قاعدة بيانات'

- Example: create

```
table instructor (
    ID          char(5),
    name        varchar(20),
    dept_name   varchar(20),
    salary      numeric(8,2))
```

➢ DDL compiler generates a set of table templates stored in a **data dictionary**. ' مترجم مجموعة من القوالب الجدول المخزنة في قاموس البيانات '

➢ Data dictionary contains metadata (i.e., data about data)

- Database schema
- Integrity constraints
  - Primary key (ID uniquely identifies instructors) . "يمنع التكرار في البيانات مثال بالجامعه: عندنا طالبتين الأسماء الرباعية الاولى تتشابه، كيف نفرق بينهم بسستم عن طريق الآي يد ستيودنت، ف تعريفه هو عمود بجدول البيانات مستحيل يتكرر مثل أي دي ستيودنت و ارقام التلفونات و ارقام البنوك "
  - Referential integrity (references constraint in SQL)
    - ✓ e.g. dept_name value in any instructor tuple must appear in department relation
  - Authorization.

➢ DDL specify a database schema by a set of definitions. It is used to specify additional properties of the data. ' تحديد سكيما قاعدة بيانات من خلال مجموعة من التعاريف. يتم استخدامه لتحديد خصائص إضافية للبيانات.

❖ **SQL**

➢ SQL widely used procedural language

- Example: Find the name of the instructor with ID 22222

```
select  name
from    instructor
where   instructor.ID = '22222'
```

- Example: Find the ID and building of instructors in the Physics dept.

```
select instructor.ID, department.buildingfrom instructor,
departmentwhere instructor.dept_name = department.dept_name and
        department.dept_name = 'Physics'
```

➢ Application programs generally access databases through one of

- Language extensions to allow embedded SQL
- Application program interface (e.g., ODBC/JDBC) which allow SQL queries to be sent to a database

❖ **Database Design**

➢ The process of designing the general structure of the database: ' عملية تصميم الهيكل العام لقاعدة البيانات:'

SHAHAD ALHAWASHI

- Logical Design – Deciding on the database schema. Database design requires that we find a “good” collection of relation schemas. تصميم '. قاعدة البيانات يتطلب أن نجد "قود" مجموعة من علاقة السكيما'
  - Business decision – What attributes should we record in the database?. ' ما سمات يجب أن تسجل في قاعدة البيانات؟ '
  - Computer Science decision – What relation schemas should we have and how should the attributes be distributed among the various relation schemas?. ' ما العلاقة السكيماز وكيف ينبغي أن توزع السمات بين مختلف علاقة السكيماز؟'
- Physical Design – Deciding on the physical layout of the database. ' الشكل الهندسي للقاعدة. '

❖ **IS THERE ANY PROBLEMS WITH THE THIS DESIGN?**

ID	*name*	*salary*	*dept_name*	*building*	*budget*
22222	Einstein	95000	Physics	Watson	70000
12121	Wu	90000	Finance	Painter	120000
32343	El Said	60000	History	Painter	50000
45565	Katz	75000	Comp. Sci.	Taylor	100000
98345	Kim	80000	Elec. Eng.	Taylor	85000
76766	Crick	72000	Biology	Watson	90000
10101	Srinivasan	65000	Comp. Sci.	Taylor	100000
58583	Califieri	62000	History	Painter	50000
83821	Brandt	92000	Comp. Sci	Taylor	100000
15151	Mozart	40000	Music	Packard	80000
33456	Gold	87000	Physics	Watson	70000
76543	Singh	80000	Finance	Painter	120000

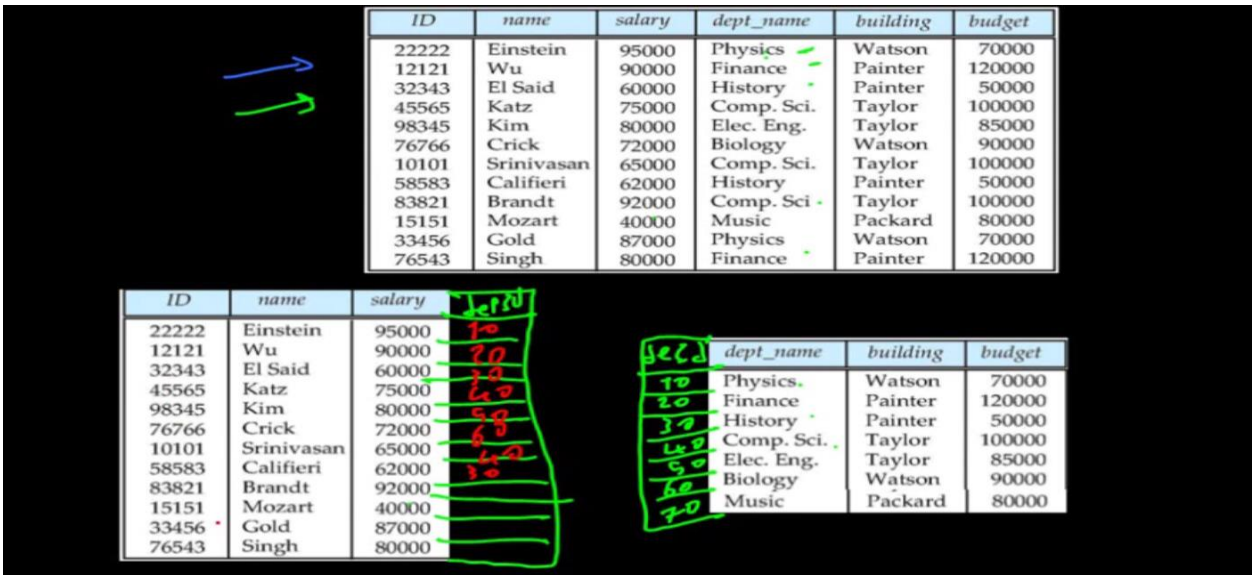
Ans. Yes

1- **خطأ بسبب تكرار البيانات في dept_name and building**

2- **الحل: نفصل الجدولين مع مراعاة عدم تكرار البيانات بالجدول رقم2**

3- **نضيف عمود بكلا الجدولين باسم dept ID ونرقم الجدول جهة اليمين**

4- **جدول جهة اليسار نرقمه بأرقام ال dept_ID على حسب علاقة dept_name & name كتالي:**



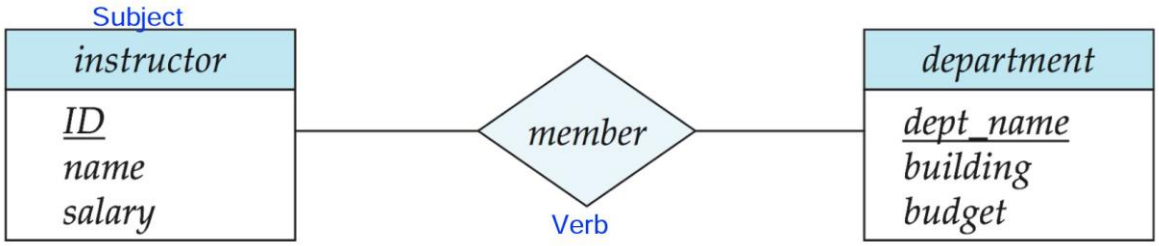
SHAHAD ALHAWASHI

❖ **The Entity-Relationship Model. Reading**

➢ Models an enterprise as a collection of entities and relationships

- Entity: a "thing" or "object" in the enterprise that is distinguishable from other objects
  - Described by a set of attributes
- Relationship: an association among several entities

➢ Represented diagrammatically by an entity-relationship diagram:



What happened to dept_name of instructor and student?.

❖ **The data models can be classified into four different categories:** قراءه

1) **Relational model:** uses a collection of tables to represent both data and the relationships among those data. Each table has multiple columns, and each column has a unique name. Tables are also known as relations.
2) **Entity-Relationship data model** (mainly for database design): uses a collection of basic objects, called entities, and relationships among these objects.
3) **Object-based data models:**
   • object-oriented model (by Java, C++, or C#) the development of an object-oriented data model that can be seen as extending the E-R model with notions of encapsulation, method (functions), and object identity.
4) **Semistructured Data Model.** The semistructured data model permits the specification of data where individual data items of the same type may have different sets of attributes.
   - **Extensible Markup Language (XML)**
   - **network data model and hierarchical data model**

Object-Relational Data Models: قرأءه غير مهم

➢ Relational model: flat, "atomic" values

➢ Object Relational Data Models

- Extend the relational data model by including object orientation and constructs to deal with added data types.
- Allow attributes of tuples to have complex types, including nonatomic values such as nested relations.
- Preserve relational foundations, in particular the declarative access to data, while extending modeling power.

SHAHAD ALHAWASHI

- Provide upward compatibility with existing relational languages.

❖ **XML: Extensible Markup Language** **قرأءه غير مهمه**

- Defined by the WWW Consortium (W3C)
- Originally intended as a document markup language not a database language
- The ability to specify new tags, and to create nested tag structures made XML a great way to exchange data, not just documents
- XML has become the basis for all new generation data interchange formats.
- A wide variety of tools is available for parsing, browsing and querying XML documents/data

❖ **Storage Management important**

- **Storage manager** is a program module that provides the interface between the low-level data stored in the database and the application programs and queries submitted to the system. ' هو وحدة البرنامج الذي يوفر واجهة بين البيانات على مستوى منخفض المخزنة في قاعدة البيانات وبرامج التطبيقات والاستفسارات المقدمة إلى النظام. '
- The storage manager is responsible to the following tasks:
  - Interaction with the file manager ' التفاعل مع مانجر ملف '
  - Efficient storing, retrieving and updating of data. ' كفاءة تخزين واسترجاع وتحديث البيانات '
- Issues:
  - Storage access
  - File organization. ' تنظيم الملفات '
  - Indexing and hashing. ' الفهرسة والتجزئة '

❖ **Query Processing** ' معالجة الاستعلام '

1.Parsing and translation ' تحليل وترجمة '
2.Optimization ' تهيئة '
3.Evaluation 'تقييم'

- Alternative ways of evaluating a given query ' طرق بديلة لتقييم استفسار معين '
  - Equivalent expressions
  - Different algorithms for each operation
- Cost difference between a good and a bad way of evaluating a query can be enormous
- Need to estimate the cost of operations ' تحتاج إلى تقدير تكلفة العمليات '
  - Depends critically on statistical information about relations which the database must maintain ' يعتمد بشكل كبير على المعلومات الإحصائية حول العلاقات التي يجب أن تحافظ على قاعدة البيانات '
  - Need to estimate statistics for intermediate results to compute cost of complex expressions ' تحتاج إلى تقدير إحصاءات عن نتائج المتوسطة لحساب تكلفة التعبيرات المعقدة '

❖ **Transaction Management important**

- What if the system fails?
- What if more than one user is concurrently updating the same data?
- A transaction is a collection of operations that performs a single logical function in a database application ' هي عبارة عن مجموعة من العمليات التي تؤدي وظيفة منطقية واحدة في تطبيق قاعدة بيانات '

SHAHAD ALHAWASHI

- **Transaction-management** component ensures that the database remains in a consistent (correct) state despite system failures (e.g., power failures and operating system crashes) and transaction failures. ' عنصر يضمن أن قاعدة البيانات لا تزال (صحيحه) حالة متناسقة على الرغم من فشل النظام (على سبيل المثال، انقطاع التيار الكهربائي وتعطل نظام التشغيل) وفشل الأجرأء. '
- **Concurrency-control manager** controls the interaction among the concurrent transactions, to ensure the consistency of the database. ' تسيطر على التفاعل بين العمليات المتزامنة، لضمان الاتساق في قاعدة البيانات. '

❖ **Briefly talking about Database Architecture?**

The architecture of a database systems is greatly influenced by the underlying computer system on which the database is running: 'يتأثر إلى حد كبير من قبل نظام الكمبيوتر الأساسي الذي تشغيل قاعدة البيانات'

1) Centralized
2) Client-server
3) Parallel (multi-processor)
4) Distributed