Key Definitions

- The **data storage function** manages how data is stored and handled by programs that run the system,
- Goals of data storage design
  - Efficient data retrieval (good response time)
  - Access to the information users need

- ال **وظيفة تخزين البيانات** تدير كيف يتم تخزين البيانات والتعامل معها من قبل البرامج التي تعمل على النظام،
- أهداف تصميم تخزين البيانات
  - استرجاع البيانات بكفاءة (جيد زمن الاستجابة)
  - الوصول إلى المستخدمين المعلومات التي يحتاج إليها

Types of Data Storage Formats:

1. **Files:** electronic lists of data optimized to perform a particular transaction
2. **Database**: a collection of groupings of information the relate to each other in some way.

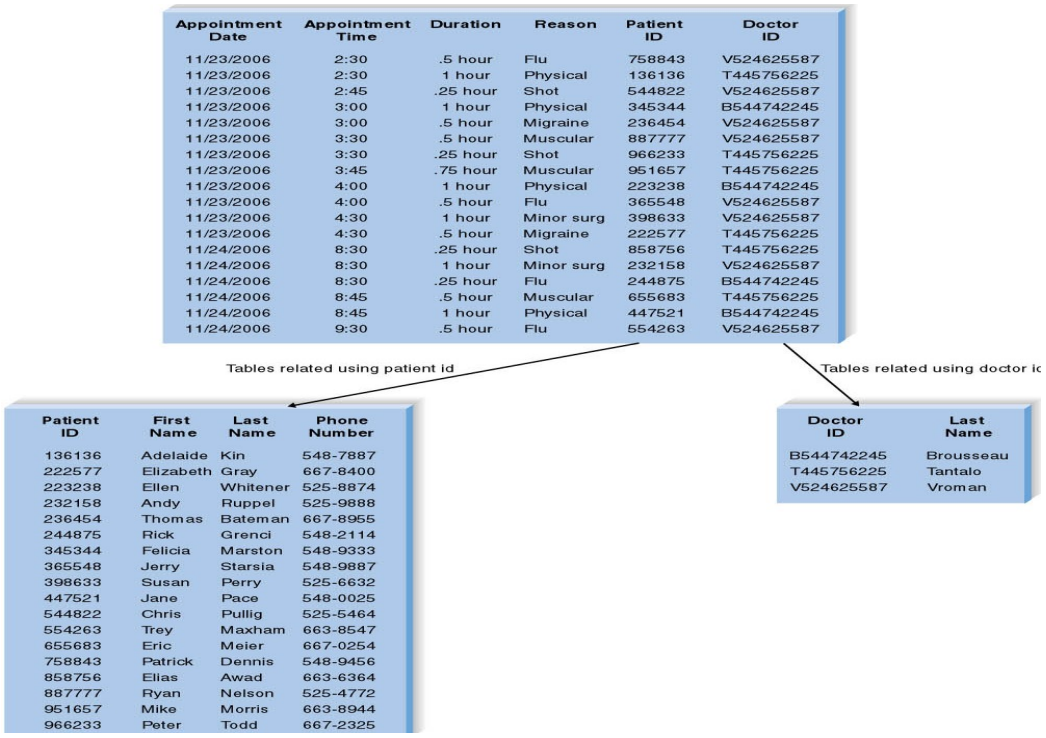
- A **Database Management System (DBMS)** is software that creates and manipulates databases.

- **ملفات :**قوائم إلكترونية من البيانات الأمثل لإجراء معاملة معينة
- **قاعدة البيانات** :مجموعة من التجمعات من المعلومات تتصل بعضها البعض بطريقة أو بأخرى.
- ا **نظام إدارة قواعد البيانات (DBMS)** هو البرنامج الذي يخلق وتعالج قواعد البيانات.

Appointment File

Appointment Date	Appointment Time	Duration	Reason	Patient ID	First Name	Last Name	Phone Number	Doctor ID	Doctor Last Name
11/23/2006	2:30	.25 hour	Flu	758843	Patrick	Dennis	548-9456	V524625587	Vroman
11/23/2006	2:30	1 hour	Physical	136136	Adelaide	Kin	548-7887	T445756225	Tantalo
11/23/2006	2:45	.25 hour	Shot	544822	Chris	Pullig	525-5464	V524625587	Vroman
11/23/2006	3:00	1 hour	Physical	345344	Felicia	Marston	548-9333	B544742245	Brousseau
11/23/2006	3:00	.5 hour	Migraine	236454	Thomas	Bateman	667-8955	V524625587	Vroman
11/23/2006	3:30	.5 hour	Muscular	887777	Ryan	Nelson	525-4772	V524625587	Vroman
11/23/2006	3:30	.25 hour	Shot	966233	Peter	Todd	667-2325	T445756225	Tantalo
11/23/2006	3:45	.75 hour	Muscular	951657	Mike	Morris	663-8944	T445756225	Tantalo
11/23/2006	4:00	1 hour	Physical	223238	Ellen	Whitener	525-8874	B544742245	Brousseau
11/23/2006	4:00	.5 hour	Flu	365548	Jerry	Starsia	548-9887	V524625587	Vroman
11/23/2006	4:30	1 hour	Minor surg	398633	Susan	Perry	525-6632	V524625587	Vroman
11/23/2006	4:30	.5 hour	Migraine	222577	Elizabeth	Gray	667-8400	T445756225	Tantalo
11/24/2006	8:30	.25 hour	Shot	858756	Elias	Awad	663-6364	T445756225	Tantalo
11/24/2006	8:30	1 hour	Minor surg	232158	Andy	Ruppel	525-9888	V524625587	Vroman
11/24/2006	8:30	.25 hour	Flu	244875	Rick	Grenci	548-2114	B544742245	Brousseau
11/24/2006	8:45	.5 hour	Muscular	655683	Eric	Meier	667-0254	T445756225	Tantalo
11/24/2006	8:45	1 hour	Physical	447521	Jane	Pace	548-0025	B544742245	Brousseau
11/24/2006	9:30	.5 hour	Flu	554263	Trey	Maxham	663-8547	V524625587	Vroman

Appointment Database



File Attributes

- **Data File** contains electronic list of information formatted for a particular transaction
- Typically organized **sequentially**
- **Pointers** - information about location of record - used to associate records with other records
- Sometimes files are called **Linked Lists** - are files with records linked together using pointers

سمات الملف:

ملف البيانات يحتوي على قائمة المعلومات الالكترونية مهيأ لمعاملة معينة

- نظمت عادة **بالتعاقب**
- **مؤشرات** - معلومات عن موقع السجل - تستخدم لربط السجلات مع سجلات أخرى
- أحيانا تسمى ملفات **قوائم مرتبطة** - ملفات مع سجلات مرتبطة معا باستخدام مؤشرات

File Types

1. **Master files** - store core, important information/ e.g. order information or customer mailing
- **Look-up files** - store static values/ e.g. list of valid codes or the names of the U.S. states
- **Transaction files** - store information that updates a master file/ Customer address till it's updated in master file
1. **Audit files** - record before and after versions of data
2. **History (archive) files** - store past information, no longer used.

١. **الملفات الرئيسية** - مخزن الأساسية، معلومات هامة / على سبيل المثال طلب المعلومات أو بريدية العملاء

- **ننظر متابعة ملفات** - قيم ثابتة متجر / على سبيل المثال قائمة رموز صالحة أو أسماء ولايات أمريكية
- **ملفات عملية** - تخزين المعلومات يقوم بتحديث ملف رئيسي / عنوان العميل حتى انها تحديثها في الملف الرئيسي

١. **ملفات التدقيق** - سجل قبل وبعد إصدارات البيانات

٢. **ملفات التاريخ(أرشيف)** - مخزن المعلومات الماضية، لم تعد تستخدم.

Database Types

1. **Legacy database**
    A. Hierarchical (depict parent-child relationships using inverted trees, 1:M relationship)
    B. Network (depict nonhierarchical associations using pointers, M:M relationship)
2. **Relational database**
3. **Object database**
4. **Multidimensional database**

١ .قاعدة بيانات تراث

A. الهرمية (تصوير العلاقات بين الوالدين والطفل استخدام الأشجار مقلوب، 1: M العلاقة)

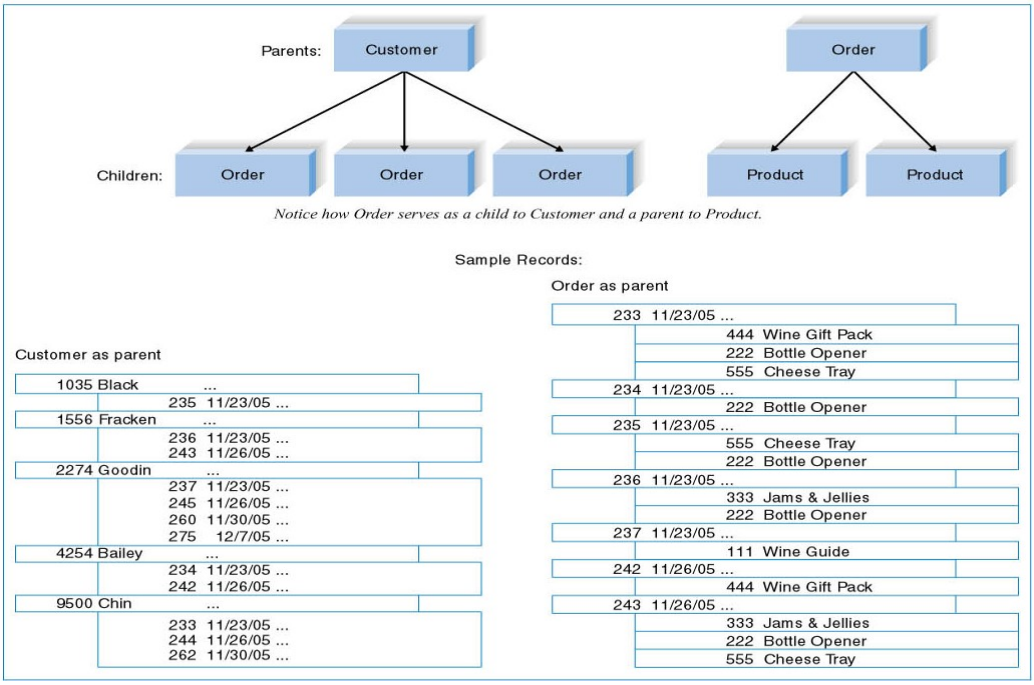
B. شبكة(تصوير الجمعيات nonhierarchical باستخدام المؤشرات العلاقةM: M)

٢ .قاعدة بيانات علائقية

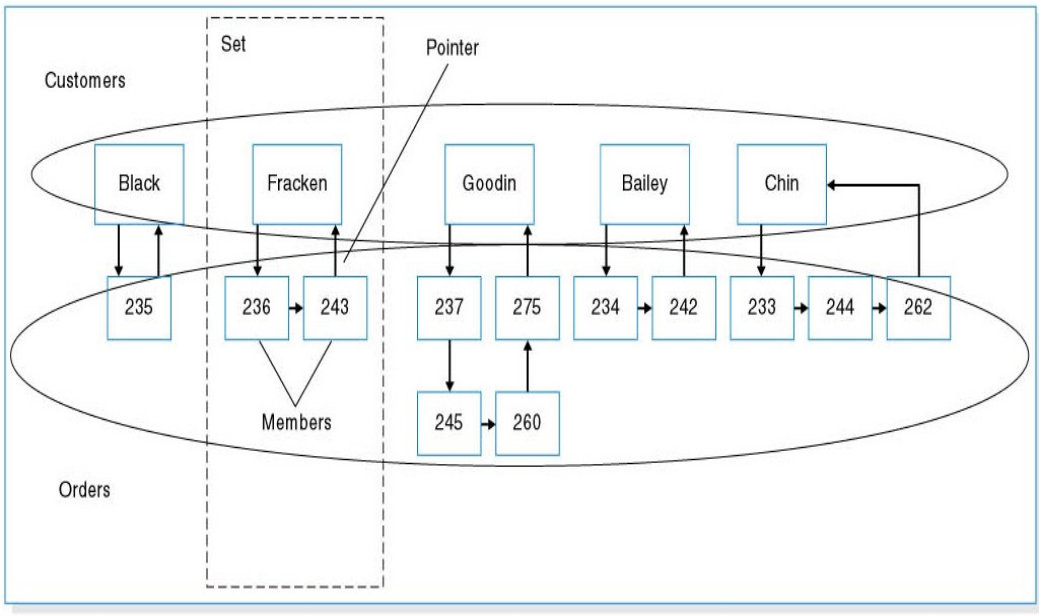
٤ .كائن قاعدة بيانات

٥ .قاعدة بيانات متعددة الأبعاد

Hierarchical Database Example



Network Database Example



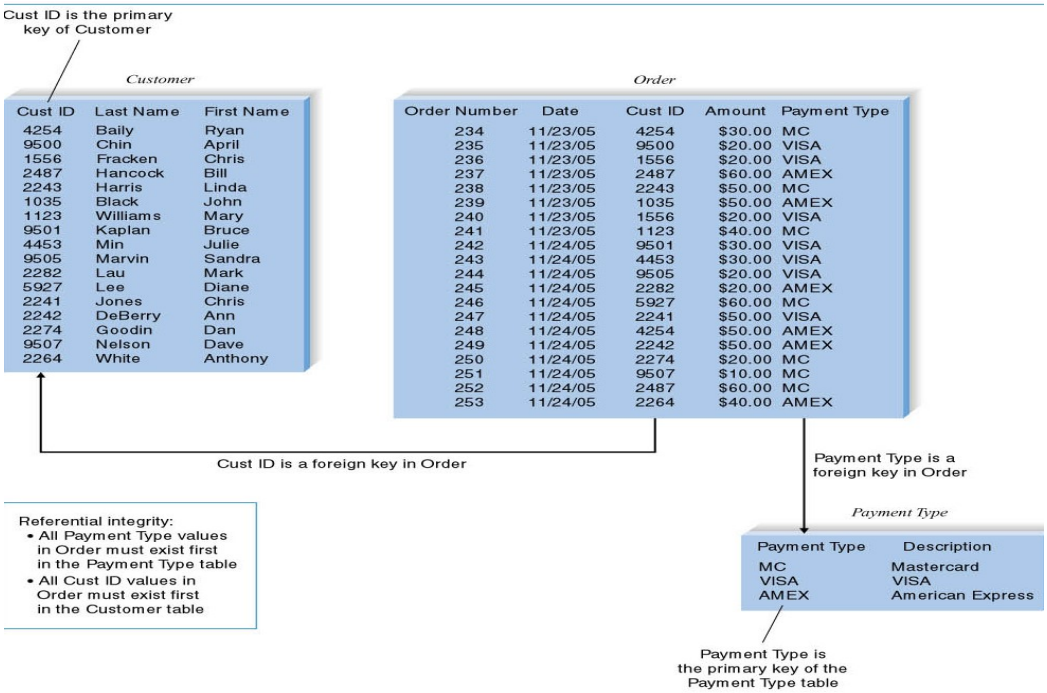
Relational Database Concepts

- Popular; easy for developers to use
- **Primary** and **foreign keys** used to identify and link tables
- **Referential integrity** ensures correct and valid table synchronization
- Structured Query Language (**SQL**)- standard language for accessing data in tables. Operate on complete table, if 2 tables are required we join them then make the query.

قاعدة بيانات المفاهيم ذات العلاقة

- جمع؛ من السهل للمطورين لاستخدام
- **ابتدائي** و **مفاتيح خارجية** تستخدم لتحديد والجداول صلة
- **التكامل المرجعي** يضمن تزامن الجدول الصحيح وصالحة
- لغة الاستعلام الهيكلية - (**SQL**) اللغة القياسية للوصول إلى البيانات في الجداول .تعمل على الجدول الكامل، إذا 2 الجداول المطلوب منا الانضمام إليهم ثم جعل الاستعلام.

Relational Database Example



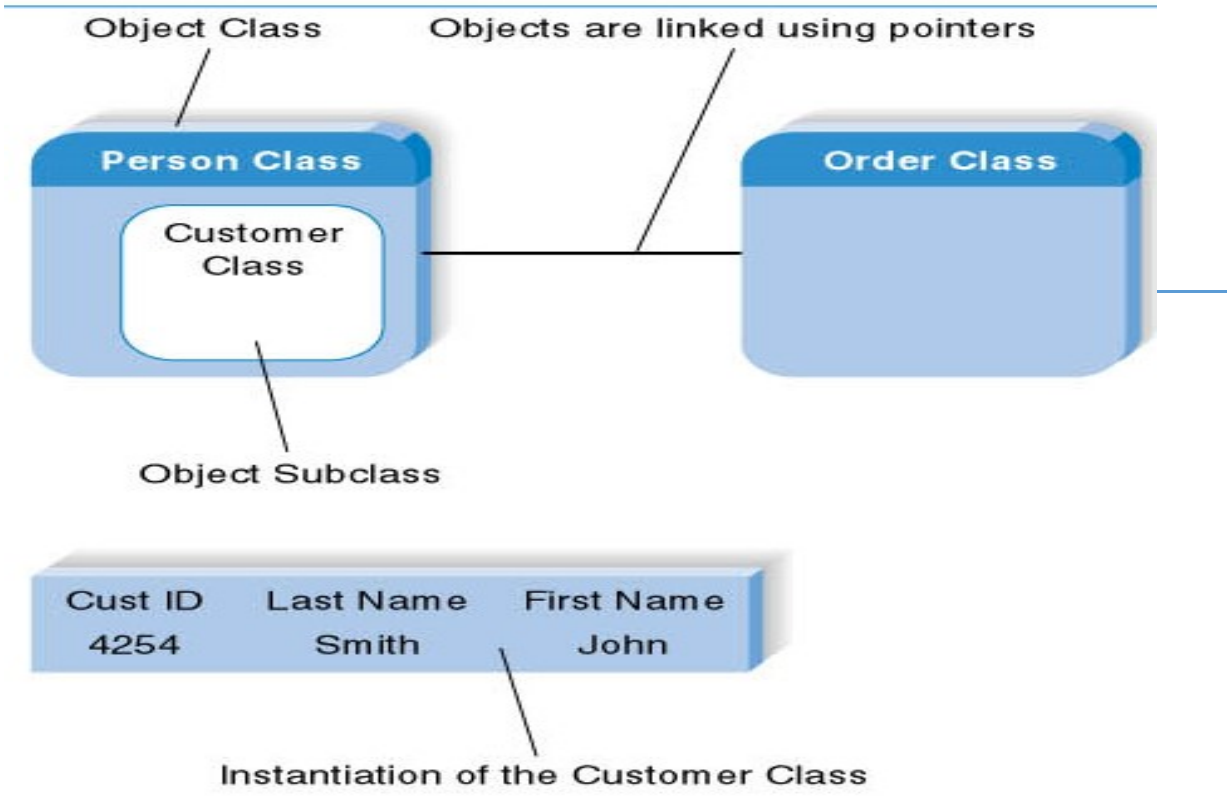
Object Database Concepts

- Built around **objects** consisting of both data and processes
- Objects are **encapsulated** (self-contained)/ means Changes to one object have no effect on other objects and allows objects to be reused to build many different systems,
- Object **classes** - major object categories contains data and processes./ may contain subclasses.
- **OODBMS** - used primarily for applications with multimedia or complex data
- **Hybrid OODBMS** - both object and relational features

قاعدة بيانات المفاهيم الكائن

- بنيت حول **شيء** تتألف من البيانات والعمليات
- الكائنات (**مغلفة** بذاته) يعني تغييرات في كائن واحد ليس لها أي تأثير على الكائنات الأخرى، ويسمح الكائنات التي سيتم استخدامها لبناء العديد من أنظمة مختلفة،
- موضوع **الطبقات** - فئات الكائن رئيسية تحتوي على بيانات وقد يحتوي /.processes فرعية.
- **OODBMS** - تستخدم في المقام الأول لتطبيقات الوسائط المتعددة مع أو البيانات المعقدة
- **OODBMS الهجينة** - على حد سواء وجوه وملامح العلائقية

object Database Example



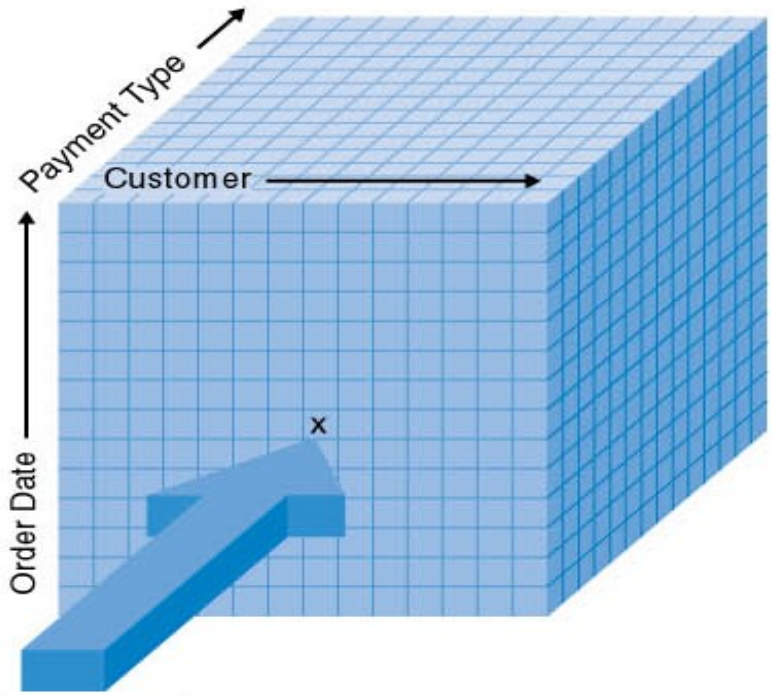
Multidimensional Database Concepts

- Stores data for easy aggregation and manipulation across many dimensions
- Used for data warehouses and data marts,
- *Data warehousing is* taking data from a company's transaction processing systems, transforming it (e.g., cleaning them up, totaling them, aggregating them), and then storing it in a data warehouse (i.e., a large database) that supports *decision support systems (DSS)*
- data marts smaller databases based on data warehouse data. That receives downloads of data from the data warehouse regularly, and it supports DSS for a specific department or functional area of the company.
- *(DSS) are designed* not to search for a particular record, but to display information that is *aggregated (e.g., totaled or averaged) across many records*
- Summary data is pre-calculated and stored for fast access

قاعدة بيانات المفاهيم متعددة الأبعاد:

- البيانات مخازن لتجميع سهلة والتلاعب عبر العديد من الأبعاد
- يستعمل ل مستودعات البيانات و سوق البيانات،
- *تخزين البيانات* هو أخذ البيانات من أنظمة معالجة المعاملات الشركة، وتحويلها) على سبيل المثال، تطهيرها، بلغ مجموعها لهم، اجمعهم(، ومن ثم تخزينها في مستودع البيانات) أي قاعدة بيانات كبيرة (التي تدعم *نظم دعم القرار (DSS)*
- سوق البيانات قواعد بيانات أصغر استنادا إلى بيانات مستودع البيانات .الذي يتلقى تنزيل البيانات من مستودع البيانات بانتظام، وأنها تدعم DSS لقسم معين أو مجال وظيفي للشركة.
- *(DSS)* قد تم *تصميمه* لا للبحث عن سجل معين، ولكن لعرض المعلومات التي مجمع) *على سبيل المثال، بلغ متوسط أو (عبر العديد من السجلات*
- بيانات التلخيص قبل احتساب وتخزينها للوصول السريع

Multidimensional Database Example



Last quarter, how many customers placed more than one order using an American Express card?

SELECTING A STORAGE FORMAT

	Files	Legacy DBMS	Relational DBMS	Object-Oriented DBMS	Multi-dimensional DBMS
Major strengths	Files can be designed for fast performance; good for short-term data storage.	Very mature products	Leader in the database market; can handle diverse data needs	Able to handle complex data	Configured to answer decision support questions quickly
Major weaknesses	Redundant data; data must be updated, using programs.	Not able to store data as efficiently; limited future	Cannot handle complex data	Technology is still maturing; skills are hard to find.	Highly specialized use; skills are hard to find
Data types supported	Simple	*Not recommended for new systems*	Simple	Complex (e.g., video, audio, images)	Aggregated
Types of application systems supported	Transaction processing	*Not recommended for new systems*	Transaction processing and decision making	Transaction processing	Decision making
Existing data formats	Organization dependent	Organization dependent	Organization dependent	Organization dependent	Organization dependent
Future needs	Limited future prospects	Poor future prospects	Good future prospects	Uncertain future prospects	Uncertain future prospects

Selecting a Storage Format (continued)

- **Data Types: It is the best choice for simple data storage and it can handle data effectively,**
- **Type of Application System: the best choices for these systems usually are relational databases and multidimensional databases as the formats can be configured**
- **Existing Storage Formats: Data storage format should be selected primarily on the basis of the kind of data and application system being developed.**
- **Future Needs: it is easier and less expensive to find relational database expertise than to find help with a multidimensional format**

اختيار شكل التخزين (تابع):

- **أنواع البيانات: أناتي هو أفضل خيار لتخزين البيانات بسيط ويمكن التعامل مع البيانات بشكل فعال،**
- **نوع من تطبيق نظام: أفضل الخيارات لهذه الأنظمة عادة ما تكون قواعد البيانات العلائقية وقواعد البيانات متعددة الأبعاد كما يمكن تكوين الأشكال**
- **تنسيقات التخزين الحالية: وينبغي اختيار شكل تخزين البيانات في المقام الأول على أساس هذا النوع من البيانات وتطبيق نظام قيد التطوير.**
- **الاحتياجات المستقبلية: فمن الأسهل وأقل تكلفة للعثور على الخبرات قاعدة بيانات علائقية من العثور على المساعدة في شكل متعدد الأبعاد**

Types of Data in Internet Systems

Data	Type	Use	Suggested Format
Customer information	Simple (mostly text)	Transactions	Relational
Order information	Simple (text and numbers)	Transactions	Relational
Marketing information	Both simple and complex (eventually the system will contain audio clips, video, etc.)	Transactions	Object add-on?
Information that will be exchanged with the special order system	Simple text, formatted specifically for importing into the special order system	Transactions	Transaction file
Temporary information	The Web component will likely need to hold information for temporary periods of time (e.g., the shopping cart will store order information before the order is actually placed)	Transactions	Transaction file

MOVING FROM LOGICAL TO PHYSICAL DATA MODELS

The Physical Entity Relationship Diagram: Five Steps

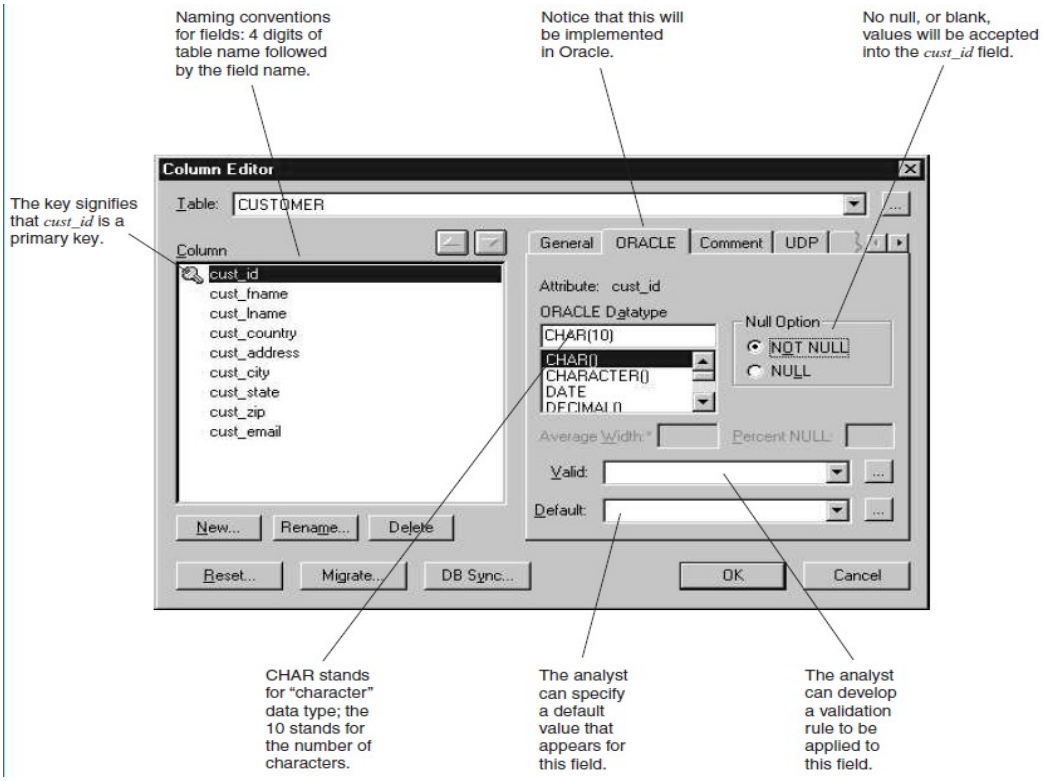
الكيان المادي مخطط العلاقة: خمس خطوات

Step	Explanation
Change entities to tables or files.	Beginning with the logical entity relationship diagram, change the entities to tables or files and update the metadata.
Change attributes to fields.	Convert the attributes to fields and update the metadata.
Add primary keys.	Assign primary keys to all entities.
Add foreign keys.	Add foreign keys to represent the relationships among entities.
Add system-related components.	Add system-related tables and fields.

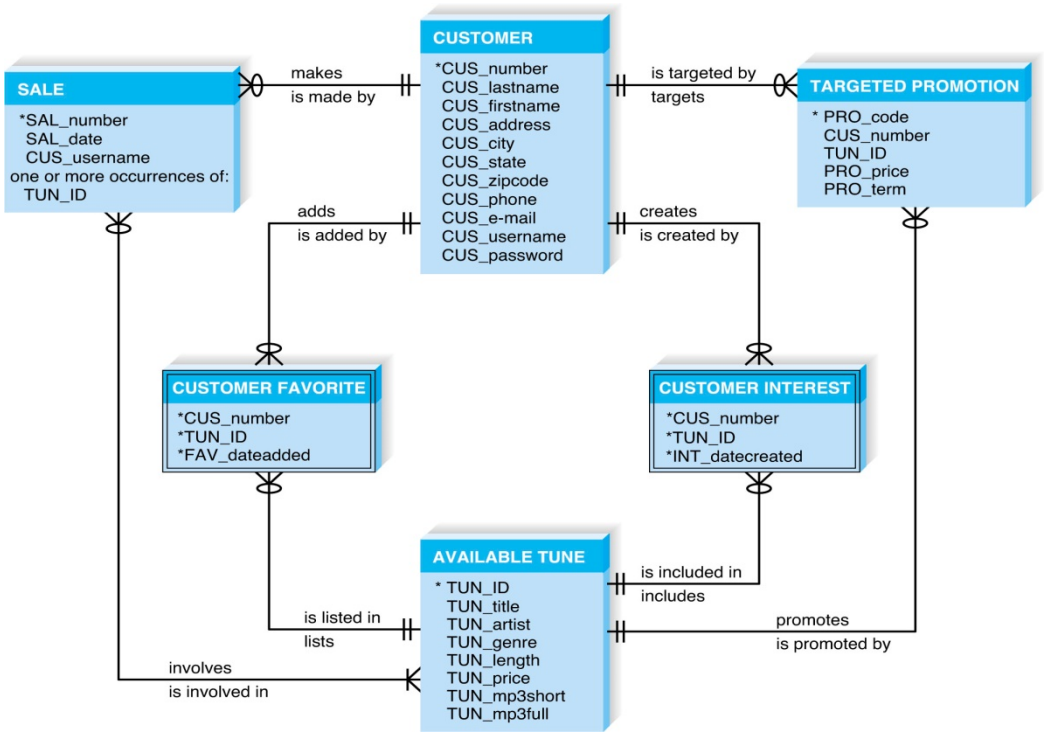
Five Step Process

- Step 1: Change Entities to Tables or Files
- Step 2: Change Attributes to Fields
- Step 3: Add Primary Keys
- Step 4: Add Foreign Keys
- Step 5: Add System-Related Components

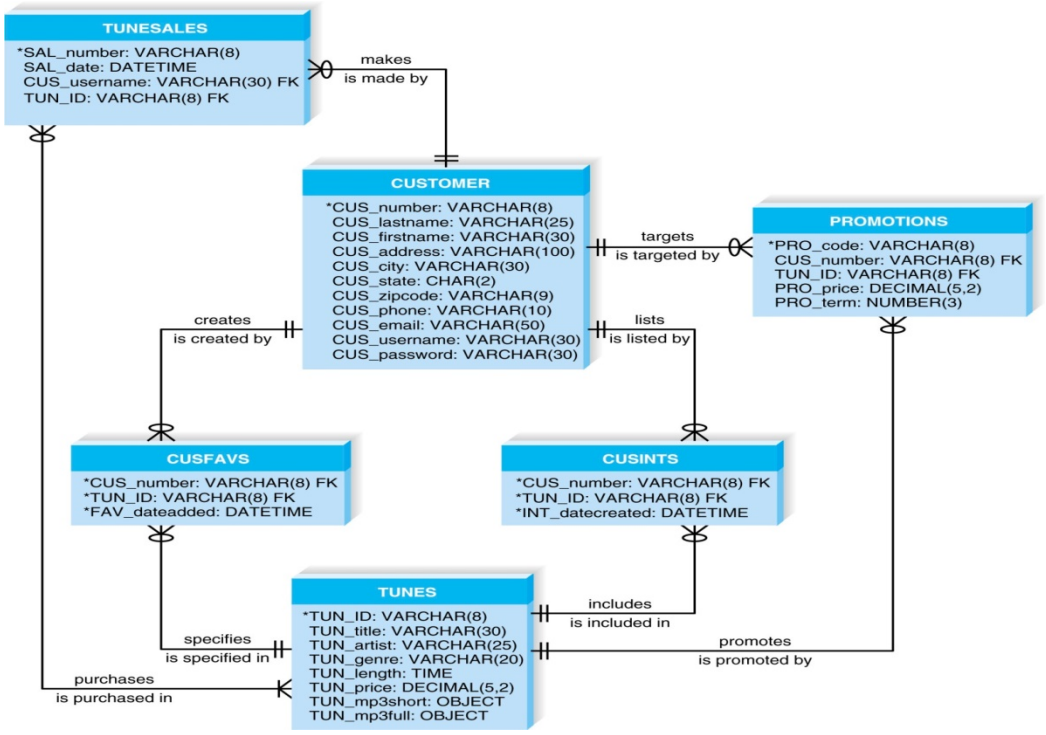
- الخطوة :1 تغيير الكيانات إلى الجداول أو الملفات
- الخطوة :2 تغيير سمات إلى الحقول
- الخطوة :3 إضافة المفاتيح الأساسية
- الخطوة :4 إضافة مفاتيح الخارجية
- خطوة :5 إضافة مكونات النظام، ذات



Tune Source logical ERD

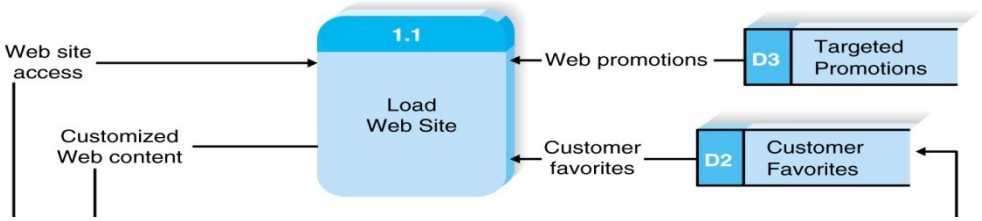


Tune Source Physical ERD



CRUD Matrix for Search and Browse Tunes Process

	1.1 Load Web Site	1.2 Process Search Requests	1.3 Process Tune Selection
PROMOTIONS			
PRO_code	R		
CUS_number	R		
TUN_ID	R		
PRO_price	R		
PRO_term	R		
CUSFAVS			
CUS_number	R		C
TUN_ID	R		C
FAV_dateadded	R		C
TUNES			
TUN_ID		R	R
TUN_title		R	R
TUN_artist		R	R
TUN_genre		R	R
TUN_length		R	R
TUN_price		R	R
TUN_mp3short		R	R
TUN_mp3full		R	R
CUSINTS			
CUS_number			C
TUN_ID			C
INT_datecreated			C



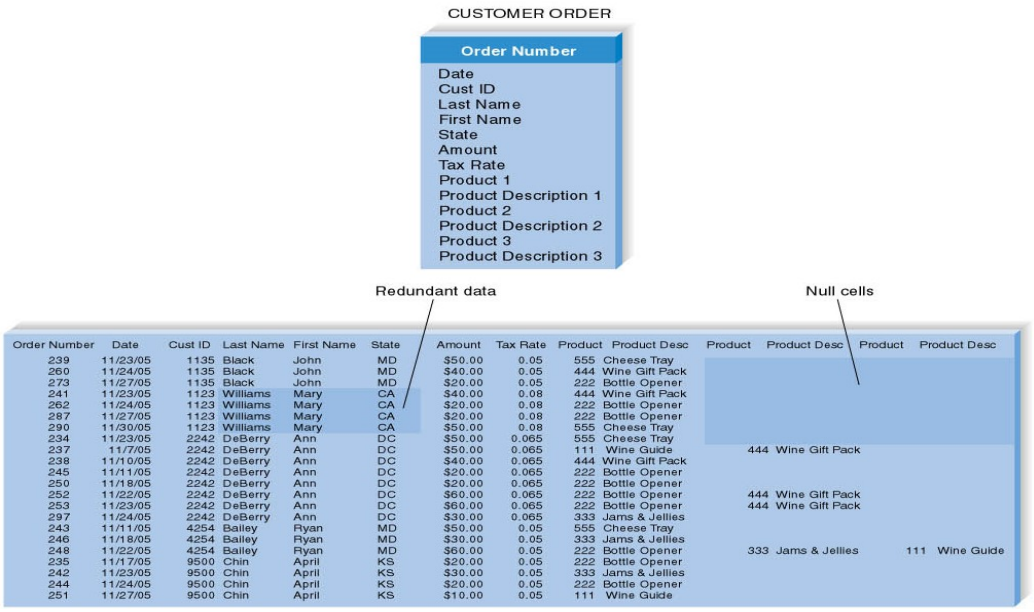
OPTIMIZING DATA STORAGE

Optimizing Storage Efficiency

- redundancy and null values should be removed from data storage design because:
- waste space, cause errors and problems integrity
- It is impossible to be sure of the actual meaning of the null values.
- Storage Efficiency done using **normalization** process

تحسين كفاءة التخزين:

- يجب إزالة القيم التكرار وخالية من تصميم تخزين البيانات للأسباب التالية :
- مساحة النفايات، يسبب الأخطاء والمشاكل سلامة
- فمن المستحيل التأكد من المعنى الحقيقي للقيم فارغة.
- تخزين الكفاءة القيام به باستخدام **تطبيع** معالج



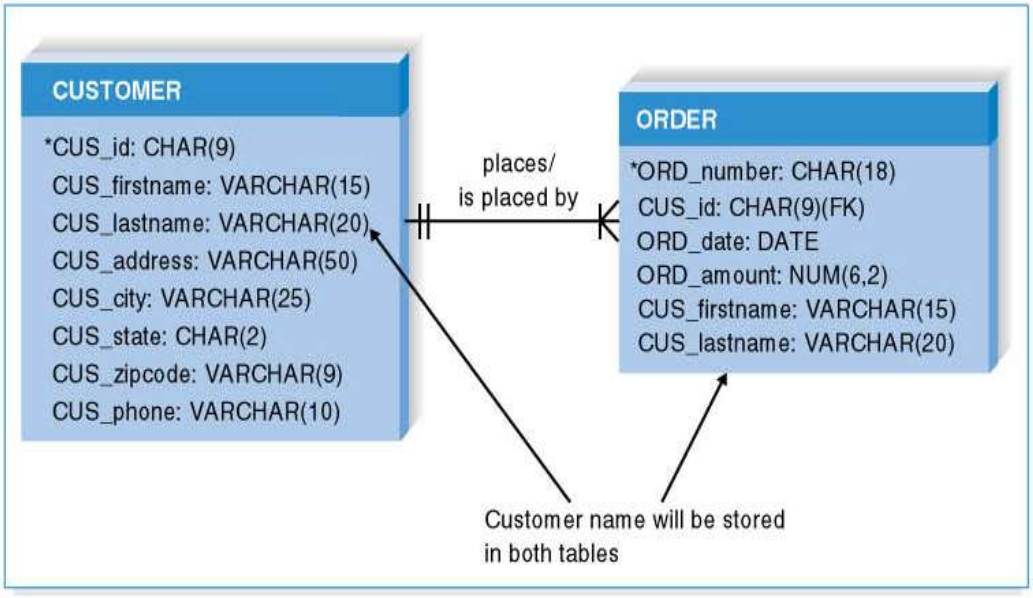
Order Number	Date	Cust ID	Last Name	First Name	State	Amount	Tax Rate	Product	Product Desc	Product	Product Desc	Product	Product Desc
239	11/23/05	1135	Black	John	MD	$50.00	0.05	555	Cheese Tray				
260	11/24/05	1135	Black	John	MD	$40.00	0.05	444	Wine Gift Pack				
273	11/27/05	1135	Black	John	MD	$20.00	0.05	222	Bottle Opener				
241	11/23/05	1123	Williams	Mary	CA	$40.00	0.08	444	Wine Gift Pack				
262	11/24/05	1123	Williams	Mary	CA	$20.00	0.08	222	Bottle Opener				
287	11/27/05	1123	Williams	Mary	CA	$20.00	0.08	222	Bottle Opener				
290	11/30/05	1123	Williams	Mary	CA	$50.00	0.08	555	Cheese Tray				
234	11/23/05	2242	DeBerry	Ann	DC	$50.00	0.065	555	Cheese Tray				
237	11/7/05	2242	DeBerry	Ann	DC	$50.00	0.065	111	Wine Guide	444	Wine Gift Pack		
238	11/10/05	2242	DeBerry	Ann	DC	$40.00	0.065	444	Wine Gift Pack				
245	11/11/05	2242	DeBerry	Ann	DC	$20.00	0.065	222	Bottle Opener				
250	11/18/05	2242	DeBerry	Ann	DC	$20.00	0.065	222	Bottle Opener				
252	11/22/05	2242	DeBerry	Ann	DC	$60.00	0.065	222	Bottle Opener	444	Wine Gift Pack		
253	11/23/05	2242	DeBerry	Ann	DC	$60.00	0.065	222	Bottle Opener	444	Wine Gift Pack		
297	11/24/05	2242	DeBerry	Ann	DC	$30.00	0.065	333	Jams & Jellies				
243	11/11/05	4254	Bailey	Ryan	MD	$50.00	0.05	555	Cheese Tray				
246	11/18/05	4254	Bailey	Ryan	MD	$30.00	0.05	333	Jams & Jellies				
248	11/22/05	4254	Bailey	Ryan	MD	$60.00	0.05	222	Bottle Opener	333	Jams & Jellies	111	Wine Guide
235	11/17/05	9500	Chin	April	KS	$20.00	0.05	222	Bottle Opener				
242	11/23/05	9500	Chin	April	KS	$30.00	0.05	333	Jams & Jellies				
244	11/24/05	9500	Chin	April	KS	$20.00	0.05	222	Bottle Opener				
251	11/27/05	9500	Chin	April	KS	$10.00	0.05	111	Wine Guide				

Optimizing Access Speed

- After you have optimized your data model design for data storage efficiency, the end result is data that is spread out across a number of tables
- There are several techniques that the project team can use to try to speed up access to data:
  - **Denormalization**/ add redundancy back into design, to reduces the number of joins in a query, applied in **look-up tables**- contain description rarely changed, to **1:1 relationship**, to **1:N relationship**, and when ***star schema design*** *is used.*
  - **Clustering**
  - **Indexing**
  - Estimating the size of data for hardware planning purposes

- بعد أن كنت قد الأمثل البيانات الخاصة بك تصميم نموذج للكفاءة تخزين البيانات، والنتيجة النهائية هي البيانات التي يتم تنتشر عبر عدد من الجداول
- هناك العديد من التقنيات التي فريق المشروع يمكن استخدامها لمحاولة تسريع الوصول إلى البيانات:
  - **Denormalization**/إضافة التكرار مرة أخرى في التصميم، ليقلل من عدد من ينضم في استعلام، وتطبق في **ابحث عن طاولات** -يحتوي على وصف نادرا ما تتغير، ل **1 :1 علاقة**، إلى **1:** **العلاقةN** ، وعندما***تصميم نجم مخطط*** *مستعمل*.
  - **تجمع**
  - **الفهرسة**
  - تقدير حجم البيانات لأغراض التخطيط الأجهزة

Denormalized Data Model



Clustering

- Reduce the number of times storage must be accessed by physically placing like records close together.
  - Interfile clustering - similar records in a table are stored together
  - Interfile clustering - combine records from more that one table that are typically retrieved together

تجمع

- تقليل يجب الوصول إلى عدد من تخزين مرات عن طريق وضع جسديا مثل سجلات قريبة من بعضها البعض.
  - Interfile يتم تخزين سجلات مماثلة في جدول معا - تجمع
  - تجمع - Interfile الجمع بين السجلات من جدول واحد أكثر من التي توجد عادة استرجاعها معا

order-id Date Cut_ID

443	2/2014	5

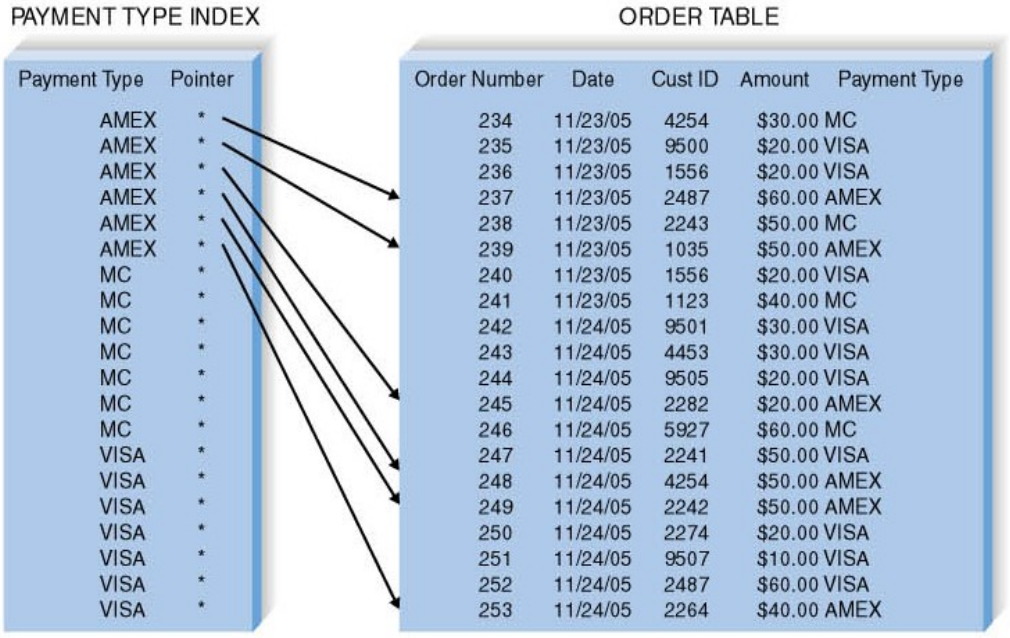
5	John	22

Indexing

- **Indexes are**: A minitable that contains values from one or more fields in a table and the location of the values within the table
- Similar to the index of a book.
- speed up the access process but slow down the update process.

- **الفهارس** A mineable :يحتوي على قيم من حقل واحد أو أكثر في الجدول وموقع القيم في الجدول
- وعلى غرار مؤشر كتاب.
- تسريع عملية الوصول ولكن إبطاء عملية التحديث.

Payment Type Index



Guidelines for Creating Indexes

- Use indexes sparingly for transaction systems.
- Use many indexes to improve response times in decision support systems.
- For each table, create a unique index that is based on the primary key.
- For each table, create an index that is based on the foreign key to improve the performance of joins.
- Create an index for fields that are used frequently for grouping, sorting, or criteria.

Estimating Storage Size
Volumetrics - Estimating Data Storage Size

- Raw data - sum of the average widths of all fields in a table.
- Calculate overhead requirements based on DBMS vendor recommendations
- Estimate initial number of records
- Estimate growth rate of records

تقدير حجم التخزين
الحجمي - تقدير حجم تخزين البيانات:

- البيانات الخام - مجموع متوسط بعرض جميع الحقول في الجدول.
- حساب متطلبات العامة بناء على توصيات بائعDBMS
- تقدير العدد الأولي من السجلات
- معدل النمو تقدير من السجلات

Calculating Volumetrics

Field	Average Size (Characters)
Order number	8
Date	7
Cust ID	4
Last name	13
First name	9
State	2
Amount	4
Tax rate	2
Record size	49
Overhead	30%
Total record size	63.7
Initial table size	50,000
Initial table volume	3,185,000
Growth rate/month	1,000
Table volume @ 3 years	5,478,200

Total record size= 30 X 49 / 100 = 14.7

14.7 + 49 = 63.7

Initial Total size (records) = 50,000

Initial Table volume=

50,000 * 63.7= 3,185,000

Growth rate/ month= 1000 ➔ in 3 years

Growth rate = 36000

Total volume @ 3 years =

(3,185,000 X 36000) / 50,000 =

2293200 + 3,185,000 = 5,478,200

Summary

- Files are electronic lists of data generally of five types: master, look-up, transaction, audit, and history.
- A database is a collection of groupings of information and a DBMS is software that creates and manipulates these.
- There are a number of methods for optimizing data access speed and data storage efficiency, though the designers may have to make tradeoffs between these goals.
- Physical ERDs contain references to how data will be stored in a file or database table.

- الملفات هي قوائم إلكترونية من البيانات عموما من خمسة أنواع :سيد، تبدو المتابعة، والمعاملات، والمراجعة، والتاريخ.
- قاعدة بيانات عبارة عن مجموعة من التجمعات المعلومات ونظم إدارة قواعد البيانات هو البرنامج الذي يخلق وتعالج هذه.
- وهناك عدد من الطرق لتحسين البيانات سرعة الوصول وكفاءة تخزين البيانات، على الرغم من أن المصممين قد تضطر إلى جعل المبادلات بين هذه الأهداف.
- ERDsالبدنية تتضمن إشارات إلى الكيفية التي سيتم بها تخزين البيانات في جدول ملف أو قاعدة البيانات.