المحور الثاني: مستودع البيانات وإدارة قاعدة البيانات (D W & DBA)

I. مدخل عام:

- ✓ Selon Vercellis (2009), l'architecture de la business intelligence contient **trois composantes majors**, on les cite comme suit : Data sources, DataWarehouse, Data marts et les méthodes de la business intelligence
- ✓ Pour comprendre le fonctionnement général des outils de *Business Intelligence*, il nous faut en amont comprendre le fonctionnement des ***datawarehouses*** et son implantation au sein du système d'information (SI) de l'organisme.
- ✓ En effet, celui-ci a longtemps occupé une place prépondérante au sein des architectures SI et prédomine toujours malgré l'émergence de nouvelles solutions.
- ✓ Dans son article sur les systèmes d'aide à la décision, publiée en 2015, Christine Sybord les décrits comme des **« systèmes [intégrant] des applications décisionnelles rassemblées dans une immense base de données, appelée « entrepôt » (Warehouse). Les données y sont thématiques**.

- ✓ Les données dites **« brutes »** car elles n'ont pas fait l'objet de traitement **prennent donc des formes différentes**, et peuvent être **structurées ou non structurées** (stockées dans leur format d'origine). **Par conséquent** : *il est impossible dans de telles dispositions de les faire communiquer **entre elles et de procéder à leur analyse***.

- ✓ **Le *Datawarehouse*** permet ainsi de palier à ce problème.

II. مفهوم مستودع البيانات (مخزن البيانات):

✓ تاريخيا:

1. **في السبعينات: ظهور قواعد البيانات الهرمية (Hierarchique database):** ترتيب البيانات في بنية تشبه بنية الشجرة (من الجذور إلى أعلى الشجرة)، أي تمثيل البيانات باستخدام **علاقات الأصل/التابع**: سمكن ان يكون *لكل أصل العديد من التوابع ولكن لكل تابع أصل واحد* (نمط العلاقات أصل واحد يردي إلى كثير من التوابع)؛ (سجل يضم مجموعة من البيانات مرتبة بالطريقة المذكورة)؛
2. **في الثمانينات: ظهور قواعد البيانات العلائقية (Relational database):** نوع من قواعد البيانات التي تخزن وتدير البيانات باستخدام **الجداول والعلاقات فيما بينها**؛
3. **في التسعينات: ظهور قواعد البيانات الموجهة نحو الهدف (الشيئية/الكائنية) (Object-Oriented databases):** التعامل مع البيانات، أثناء تنظيمها في قواعد بيانات على أساس تشبيهها بما يدور في الكائن الحي؛
4. **منذ بداية الألفية الثالثة:** ظهور مستودعات البيانات **(DW)**

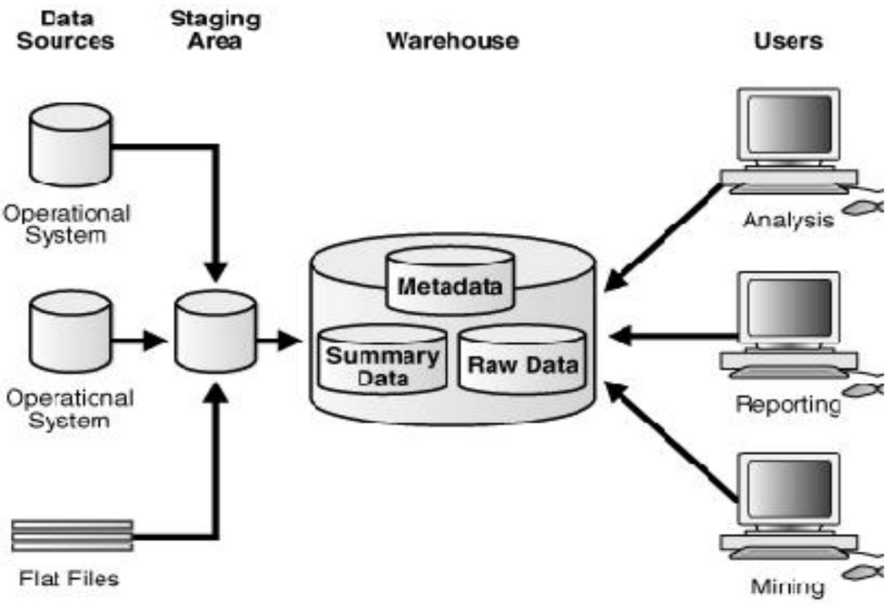
✓ مستودع البيانات (DW) هي من **أدوات ذكاء الأعمال**، وهي "عبارة عن **قاعدة بيانات** تؤمن التقارير وأدوات الاستعلام، وتقوم بتخزين البيانات الجارية والبيانات التاريخية (الجديدة والقديمة) والإحصائية، التي تستخلص من نظم العمليات المختلفة وتوحدها لأغراض الحصول على التقارير والتحاليل الإدارية وصناعة القرار"

✓ Un Entrepôt de données (DW) est une collection de données : orientées sujet(موجهة موضوع) , intégrées(متكاملة/مدمجة) , non volatile(غير متطايرة: فبمجرد وجودها في المستودع فهي تكون ثابتة) et historisées(منظمة تاريخيا) , organisées pour le support d'un processus d'aide à la décision.

✓ عامل أساسي لنجاح المنظمات في نجاح مشروع ذكاء الأعمال؛

✓ "الهدف الأساسي من تأسيس مستودع البيانات: **إنشاء عرض منطقي وحيد لبيانات المؤسسة** (يقع عادة ضمن قاعدة بيانات منفصلة ماديا)، ويستطيع المطوّرون ومتخذو القرار الوصول إليه"؛

✓ مخازن (مستودعات) البيانات هي مصدر رئيس للبيانات التي تستخدمها متاجر البيانات المتخصصة (Datamarts) يتضمن كل واحد منها بيانات خاصة بمجال معين، وبالاعتماد على تقنية الاستخراج، التحويل والتحميل (ETL) يتم خزن البيانات من عدة مصادر في مخازن البيانات ومتاجر البيانات من أجل دعم تحاليل ذكاء الأعمال؛



المصدر: صورة مأخوذة من الأنترنيت

- **Matdata**: البيانات الوصفية؛
- **Raw data**: مسودة البيانات؛
- **Summary data**: موجز البيانات؛

الفرق بين قاعدة البيانات ومستودع البيانات: الأولى مجرد مكان للتخزين، أما الثانية فتتضمن التخزين والتحليل؛

III. فوائد مستودع البيانات:

- يعمل على تجميع البيانات المتشابهة بين مختلف أقسام المنظمة (معلومات عن العملاء مثلا والتي يمكن أن نجدها في العديد من أقسام المنظمة) في قاعدة بيانات كبرى (مستودع)، تعتبر الممول الرئيسي لكل أقسام المنظمة، بحيث يتم في كل مرة تحيين المستودع؛
- يعمل على تقديم المعلومات المطورة والمحسنة، وتسهيل حصول صانع القرارات عليها؛
- تأمين القدرات والإمكانات لنمذجة البيانات، وإعادة نمذجتها؛
- تمكن صناع القرار من الوصول إلى البيانات متى احتاجوها، ومن دون التأثير على أداء نظم المعلومات الجارية ذات الأولويات؛
- تستخدم المؤسسة البيانات والمعلومات المجمعة في مستودع البيانات بطريقة المعالجة التحليلية على الخط (OLAP)، والتنقيب في البيانات (DM) ، لكي تساعدهم في إعادة تسليط الأضواء على أعمالهم؛

IV. خصائص وصفات مستودعات البيانات:

✓ تستخدم النموذج المتعدد الأبعادMultidimensional mode .

✓ تدعم السلاسل الزمنية Time series وتحليل التوجهات Trends analysis اللذين يحتاجان لبيانات تاريخية لا تستطيع قواعد البيانات العادية أن توفرها؛

✓ تحديث البيانات والذي يتم كل فترة بواسطة أجزاء من المستودع تختص بهذا الأمر.

✓ استرجاع البيانات وتحليلها وهو صميم عملها وتهتم به أكثر.

✓ مستويات تجميع Aggregation وأبعاد Dimensions غير محددة.

✓ دعم معمارية العميل/الخادم وتعددية المستخدمين.

✓ الاحتفاظ بكمية ضخمة من البيانات قد تصل إلى عدة تيرابايتات.

✓ أنها محددة لموضوع Subject-Oriented على سبيل المثال في منشأة طبية ما تكون موضوعاتها تتعلق بالأطباء، المرضى، الممرضون، الأدوية... الخ فهي أشياء حسية ومعنوية.

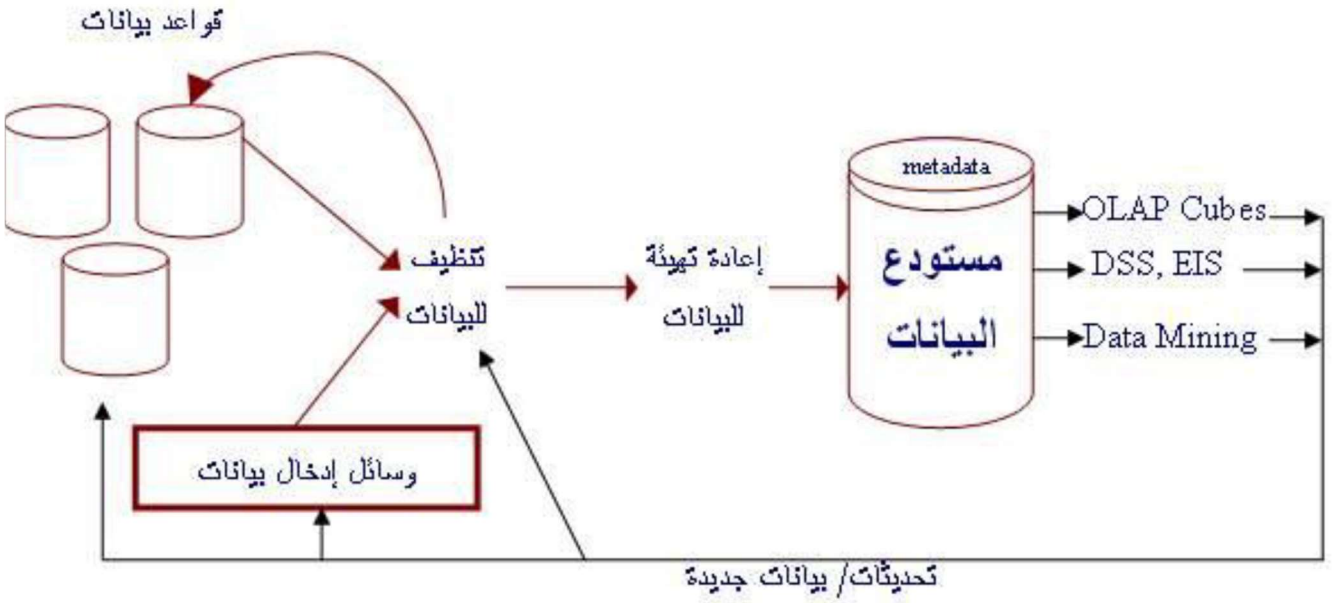
✓ أنها متكاملة Integrated من خلال وجود علاقة بين البيانات.

✓ أنها غير قابلة للتعديل Non-Volatile أن تلك البيانات عند تحميلها في المستودعات تستخدم فقط للتحليل والدراسة والعرض.

✓ أنها مرتبطة بعامل الوقت Time-Variant وهي من أهم نقطة في تلك البيانات.

V. خطوات بناء مستودعات البيانات:

شكل رقم (): بناء مستودعات البيانات



المصدر: عبد الله أحمد، 2009، مستودعات البيانات، Cybrarians Journal، العدد 18، البوابة العربية للمكتبات والمعلومات

1) **إنشاء مساحة للبيانات** Creating data preparation area وهي قاعدة بيانات ذات سعة تخزين عالية جداً تقوم بتخزين كافة البيانات القادمة من أنظمة التشغيل المختلفة **لكي يتم تنقية وتعديل البيانات فيها قبل تحميلها** في مستودع البيانات، ويراعى فيها أن يكون تصميم قاعدة إعداد البيانات متوافقاً مع تصميم مستودع البيانات.
2) **بناء مستودع البيانات** Creating data warehouse وفيها ينشأ مستودع للبيانات والذي تحمل فيه بعد استخراجها وتنقيتها، و**تصمم المستودعات دائما بحيث تسمح بوجود علاقات ذات أبعاد مختلفة.**
3) **تجزئة مستودع البيانات إلى مجموعة من متاجر البيانات Datamarts** بحيث يكون هناك بيانات خاصة بالإدارة المالية، وأخرى خاص بإدارة الموارد البشرية أو يكون التقسيم مبنياً على فروع المؤسسة.
4) **دمج وتنقية ونقل البيانات** Data extraction & cleansing & transformation وفي هذه المرحلة يتم جلب البيانات من مصادرها المختلفة إلى Data preparation area ، ونقوم بتحويلها من صورة إلى أخرى إذا تطلب الأمر ذلك، وفي أحيان كثيرة تدمج بعض البيانات مع بعضها البعض، أو نقوم بتعريف بيانات جديدة لم تكن موجودة من قبل، بالإضافة إلى تنقية البيانات غير الصحيحة وحذف الغير مهم منها.
5) **تحميل البيانات في مستودع البيانات** Loading data in data warehouse وفي هذه المرحلة تحمل البيانات من Data preparation area إلى مستودع البيانات، ويتم فيها اختبار البيانات.
6) **تحليل البيانات و إنشاء تطبيقات نظم دعم اتخاذ القرار** DSS & creating OLAP applications وفي هذه المرحلة تنفذ التطبيقات الخاصة بعرض مستودع البيانات وتحليلها وتسمى Processing systems online analytical وهذه

التطبيقات تقوم بعرض البيانات بعدة أبعاد وتقوم باستخدام خوارزميات معقدة لتحليل البيانات، أيضاً في تلك المرحلة تستخدم أدوات التنقيب عن البيانات.

VI. ما يجب مراعاته عند بناء مستودعات البيانات:

- استخلاص البيانات من عدة مصادر قد تكون غير متجانسة.
- تهيئة البيانات لضمان انسجامها Consistency داخل مستودع البيانات.
- تنظيف البيانات لضمان شرعيتها Validity ، ويتم من خلال قاعدة البيانات التي أخذت منها البيانات.
- مراقبة وضبط حجم مستودع البيانات أثناء وبعد تحميله بالبيانات.
- تحديث البيانات كل فترة من الزمن.
- تحديد الوقت اللازم للبناء وما هي الجدوى الاقتصادية بالنسبة للمؤسسة التي سوف تستعمله.

5. معظم المؤسسات تعتمد بنية **ثلاثية المستويات في بناء مستودعات البيانات**:
   - مستوى الأنظمة التشغيلية ومصادر البيانات المختلفة Data sources and operation systems .
   - مستوى مستودع البيانات Data warehouses tier.
   - مستوى تطبيقات العرض والتحليل OLAP systems.

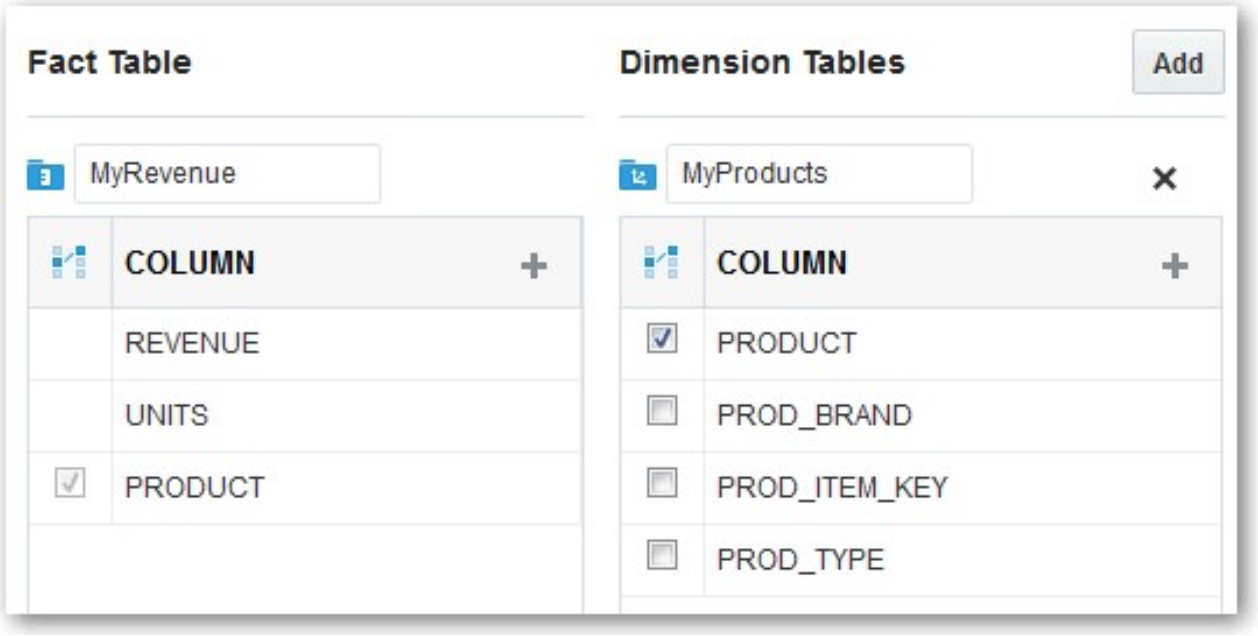
VII. كيف تخزن البيانات في مستودع البيانات؟

- ✓ تخزين البيانات يتم بطرق متعددة الاتجاهات أو ما يعرف بالنموذج متعدد الأبعاد؛
- ✓ لتخزين البيانات يجب الاستفادة من العلاقات بين البيانات لوضعها في مصفوفات متعددة الأبعاد تسمى مكعبات البيانات Data cubes؛
- ✓ إذا استعملت قواعد البيانات العلائقية في مخازن البيانات وطريقة التحليلOLAP تسمى في تلك الحالة OLAP relational، أو باستعمال قواعد بيانات متعددة الاتجاهات فتسمى Multidimensional OLAP
- ✓ عملية تخزين النموذج متعدد الأبعاد يتضمن نوعين من الجداول:
  - **جدول البعد Dimension table** وصفوفه تصف سمات البعد.
  - **جدول الحقيقة Fact table** ويتضمن مقاييس أداء العمل، وتعرف بمؤشرات لجداول الأبعاد ويتضمن هذا الجدول البيانات.

تشتمل جداول البيانات الفعلية وجداول الأبعاد على أعمدة لتخزين البيانات للنموذج:

- **تحتوي جداول البيانات الفعلية** على مقاييس، وهي عبارة عن أعمدة تشتمل على تجميعات تم إنشاؤها في التعريفات الخاصة بها. على سبيل المثال، عمود "الإيرادات" وعمود "الوحدات" هما عمودا مقاييس.
- **تحتوي جداول الأبعاد** على السمات التي تصف كيانات الأعمال. على سبيل المثال، أعمدة "اسم العميل" و"المنطقة" و"العنوان" هي أعمدة سمات.

المصدر :

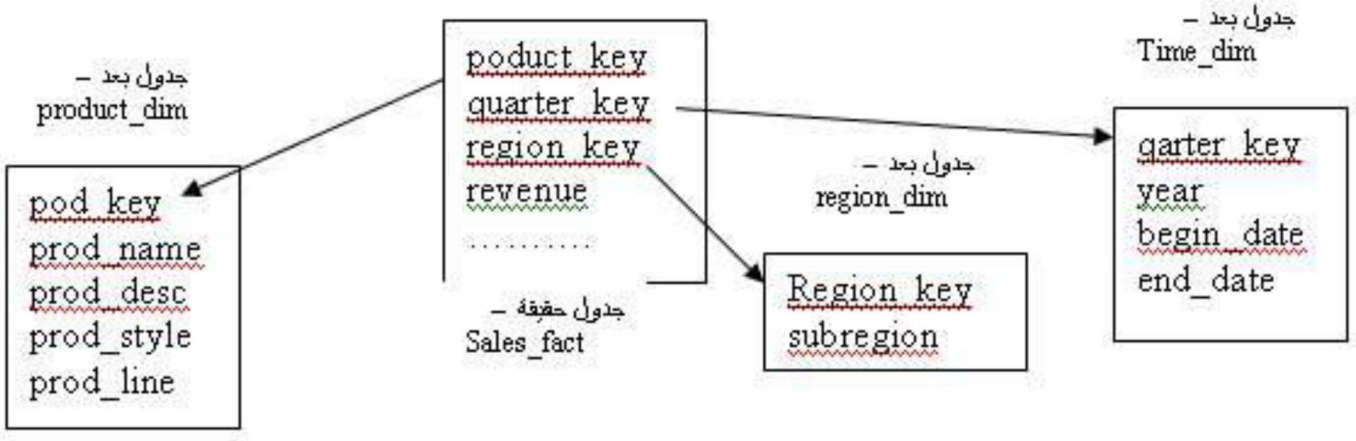


المصدر: صورة مأخوذة من الأنترنيت

ملاحظة هامة: قبل البدء في إنشاء جداول الأبعاد يجب التأكد من أن البيانات التي نحتاجها للنموذج موجودة في قائمة الجداول المصدر

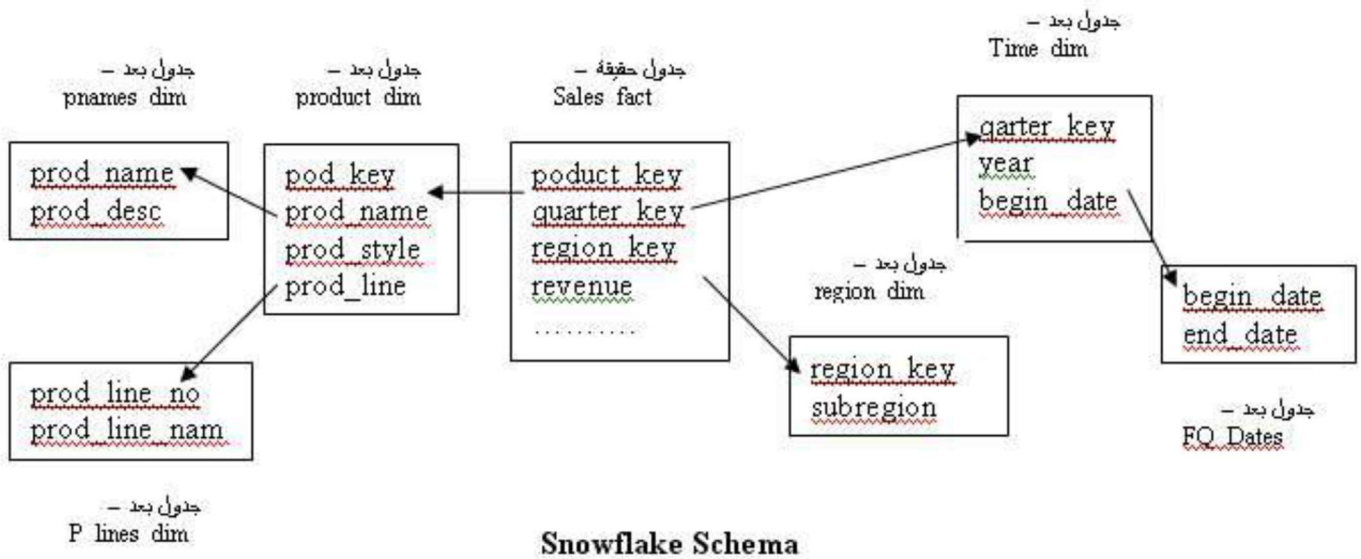
✓ توجد بنيتين شائعتين لتعريف (لتمثيل) هذه الجداول:

أولاًStar schema : (تمثيل النجمة): تتكون من جدول الحقيقة مع جدول وحيد لكل بعد، ويعيبها احتياجها لمساحة كبيرة لتخزين البيانات.



المصدر: صورة مأخوذة من الأنترنيت

ثانياً Snowflake schema : (تمثيل رقاقة الثلج): وهي تعديل للجداول الأولى Star schema وتحمل جدول حقيقة وأكثر من جدول بعد، ومشكلتها تضييعها لطاقة الجهاز المضيف في عمليات الربط.



Snowflake Schema

المصدر: صورة مأخوذة من الأنترنيت

VIII. كيف تعرض النتائج في مستودع البيانات؟

الطريقة الأولى: وهي **الحسابية** حيث تستعمل تحليل البيانات Analyze data من أجل إظهار النتائج بطريقة حسابية إحصائية (الجمع، الطرح، المعدل...الخ)؛

الطريقة الثانية: إظهار **البيانات** في شكل رسم بياني أو إحصائي Diagram بحيث تحوى كل واحدة منها معلومات في مجال معين (الإدارة، التسويق، شؤون الموظفين... الخ) لتستعمل في تحديد البحث واستعمال أنظمة التحليل على تلك المخازن لاتخاذ القرارات.

IX. نظم إدارة قواعد البيانات Database Management Systems (DBMS):

- ✓ مجموعة من البرامج التي تسمح للمستخدم بأنشاء وصيانة قواعد البيانات؛
- ✓ مجموعة من الأدوات البرمجية التي تتحكم في الوصول والتنظيم والتخزين وإدارة واسترجاع والحفاظ على البيانات في قاعدة بيانات؛
- ✓ **نظم إدارة قواعد البيانات تعتبر حلقة الوصل بين البيانات المخزنة في قاعدة البيانات وبين برمجيات التطبيق المستخدمة في مرحلة المعالجة.**
- ✓ وعليه، يتكون نظام إدارة قواعد البيانات من جزيئين:

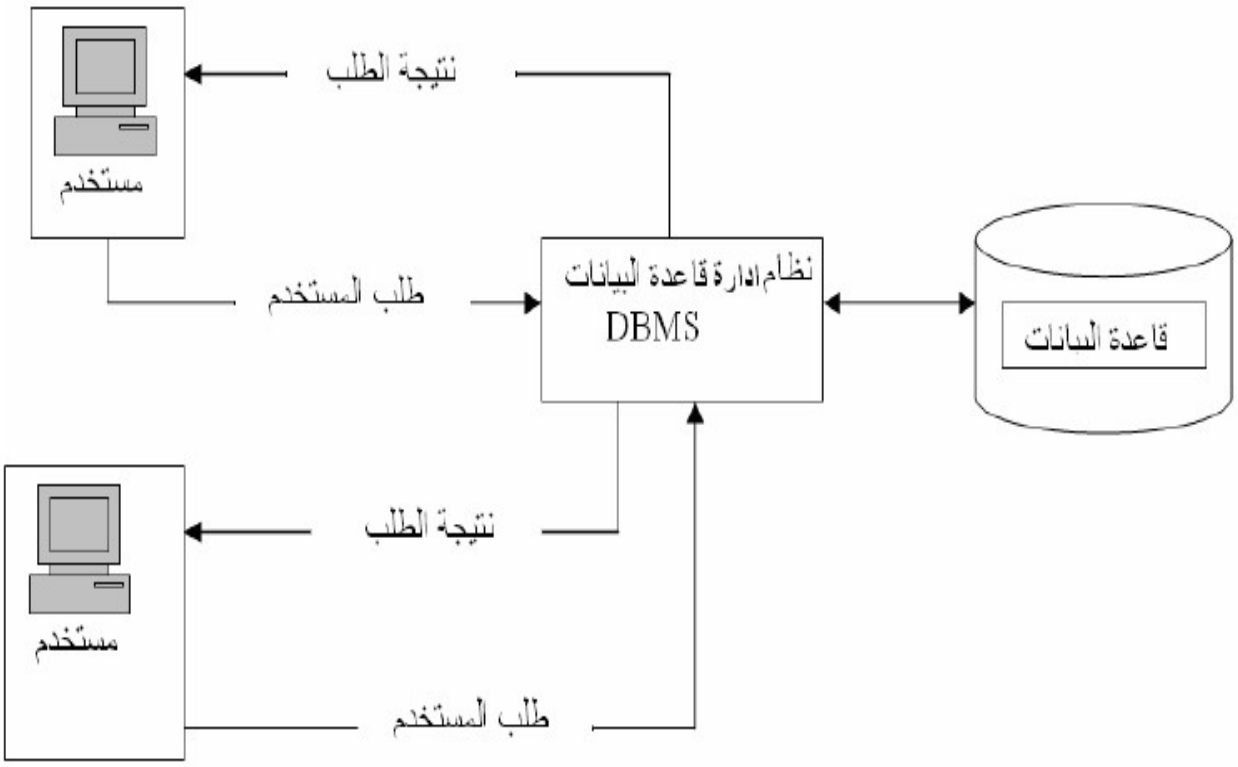
الجزء الاول: هو قلب نظام قواعد البيانات، وهو المسؤول عن إنشاء وصيانة وقراءة البيانات والتعامل معها بشتى الصور ويمكن القول بأنه محرك قاعدة البيانات Database Engine.

الجزء الثاني: هو عبارة عن مجموعة برامج وأدوات نظام تستطيع من خلالها أن تصل بمحرك قاعدة البيانات وتنفيذ الأعمال المطلوبة والمتصلة بقاعدة البيانات، وغالباً ما تكون تلك البرامج قابلة للاستدعاء من خلال نظم التشغيل أو من خلال لغات البرمجة.

✓ تميز (تفصل) نظم إدارة قواعد البيانات بين ثلاثة مناظر (صور) لقاعدة البيانات وهي:

1. **المنظر الداخلي Internal view:** هو مستوى البناء المادي للملفات، أي الكيفية التي يتم بها حفظ السجلات المكونة لإحدى الملفات على وسيط التخزين، وهناك العديد من أساليب حفظ البيانات مثل الملفات المفهرسة والملفات المباشرة؛
2. **المنظر المنطقي Logical view:** في هذا المستوى يتم توصيف الملفات وفق البنية المنطقية، حيث يتم تحديد الملفات الضرورية لبناء قاعدة البيانات وتحديد السجلات التي تتكون منها تلك الملفات، كذلك تعيين حقول السجلات ونمط البيانات التي سوف تحفظ فيها، وذلك يتم من خلال لغة توصيف البيانات Data descriptive language التي تعرف اختصاراً (DDL)، وبتعبير آخر المستوى المنطقي يمثل بناء الجداول التي تنسجم مع حاجات كافة مستخدمي النظام وليس حسب حاجة كل مستخدم بشكل مستقل؛
3. **المنظر الخارجي External view:** الذي يمثل المستخدم الخارجي أو المبرمج، حيث تتم عملية الربط بين قاعدة البيانات ومبرمجي التطبيقات من خلال لغة معالجة البيانات Data manipulation language (DML)، وبين المستخدمين وقاعدة البيانات بواسطة لغة الاستعلام Structure of Query language (SQL).

يوضح الشكل الموالي هيكلة نظام إدارة قواعد البيانات:



✓ **قاموس البيانات (DD):**

مجموعة من المعلومات عن جداول وفهارس البيانات تحفظ داخل هذا القاموس لاستعمالها من قبل نظام إدارة قواعد البيانات؛

✓ في الـ80، تم دمج النظم الفرعية للبيانات الوظائفية (نظام البيانات للمحاسبة، نظام البيانات للتسويق....) ضمن نظام واحد سمي بنظام **تخطيط موارد المؤسسة (ERP)**، وتم استبدال مخططات تمثيل البيانات القديمة بأنظمة **إدارة قواعد البيانات العلائقية (RDBMS)** **R**elational **D**ata**b**ase **M**anagement **S**ystem، مما أدى إلى تحسين الحصول على البيانات وتخزينها، فضلا عن العلاقات بين حقول البيانات التنظيمية + انخفاض ملحوظ في تكرار البيانات؛

قائمة المصادروالمراجع:

1. Ouarda Lina BENSEHAMDI,2022, L'impact du « business intelligence » sur les fonctions d'entreprises, Les ouvrages du CRASC.
2. Tom Blachon, 2022, Les outils de Business Intelligence dans le secteur des études et de la communication,

domain_shs.info.docu. 2022. mem_03847446 (HAL open science, Id: mem_03847446).

3. عبد الله أحمد، 2009، مستودعات البيانات، Cybrarians Journal، العدد 18،البوابة العربية للمكتبات والمعلومات.
4. أحسن طيار، ذكاء الأعمال ودوره في اتخاذ القرارات في البنوك، المجلة الجزائرية للدراسات المالية والمصرفية، المجلد 6، العدد 01، ديسمبر 2016.
5. Sofiane AOUAG, 2024, Cours Business Intelligence, Département Informatique Faculté des Mathématique et de l'Informatique, Université de Batna II.
6. راميش شاردا وآخرون، 2020، ذكاء الأعمال والتحليلات وعلم البيانات، منظور إداري، ترجمة محمد بن عايض القرني، مركز البحوث والدراسات، معهد الإدارة العامة، المملكة العربية السعودية.