المحاضرة الثانية: مستودع البيانات وإدارة قاعدة البيانات

1- ما هي قاعدة البيانات؟

قاعدة البيانات databaseعبارة عن مجموعة من البيانات المنظمة التي يتم تنظيمها وتخزينها بطريقة تسمح باسترجاع المعلومات وتحديثها وإدارتها بكفاءة، يتم استخدامها عادةً لمعالجة بيانات المعاملات، حيث يكون الوصول في الوقت الفعلي إلى المعلومات الحديثة أمرًا بالغ الأهمية. تُستخدم قواعد البيانات بشكل شائع في تطبيقات مثل أنظمة إدارة علاقات العملاء (CRM)، ومنصات التجارة الإلكترونية، وأنظمة إدارة الموارد البشرية[1].

تقوم قواعد البيانات بمعالجة المعاملات اليومية في المؤسسة،ومن أمثلة استخدام قاعدة البيانات مايلي:

- موقع ويب للتجارة الإلكترونية يقوم بإنشاء طلب لمنتج تم بيعه؛
- شركة طيران تستخدم نظام الحجز عبر الإنترنت؛
- مستشفى يسجل مريضا؛
- قيام البنك بإضافة معاملة سحب من ماكينة الصراف الآلي إلى الحساب.

2- ما هو مستودع البيانات؟

مستودع البيانات Data Warehouse هو مستودع بيانات يحتوي على بيانات تاريخية من مصادر غير متجانسة وهو مصمم للاستعلام والتحليل وليس لمعالجة المعاملات، بالإضافة إلى ذلك، يتكون من الأدوات والتقنيات المتاحة للاستخراج والتحويل والتحميل، ومحرك OLAP، وأدوات تحليل العميل والتطبيقات الأخرى التي تستخدم لإدارة ومعالجة البيانات لتوفير دعم القرار للعاملين في مجال المعرفة أو اتخاذ القرار[2].

أحد أسباب بناء مستودع البيانات هو القيام باستخراج البيانات. والفكرة هي البحث عن أنماط مثيرة للاهتمام ومهمة في قاعدة بيانات ضخمة.يرتبط استخراج البيانات بأنظمة اكتشاف المعرفة، والتطبيقات التي تحاول فهم البيانات. تعتبر قواعد البيانات عنصرا هاما في هذه الأنظمة لأنها تحتوي على كميات كبيرة من البيانات التي قد تزود متخذ القرار بالمعرفة. تساعد برامج استخراج البيانات واكتشاف المعرفة على التدقيق في كل هذه البيانات وتقديم الأفكار لصانع القرار[3].

أحد أهداف مستودع البيانات هو المساعدة على فهم العمل بشكل أفضل، يساعد هذا النوع من التكنولوجيا في إنشاء "منظمة تعليمية"، وهي منظمة قادرة على فهم سوقها وعملائها ونفسها بشكل أفضل[4].ومن أمثلة استخدام مستودع البيانات مايلي:

- تقسيم العملاء إلى مجموعات مختلفة بناءً على مشترياتهم السابقة لتزويدهم بمحتوى أكثر تخصيصًا؛

[1]EFRAIM TURBAN et autres, **Information Technology for Management Advancing Sustainable, Profitable Business Growth**, John Wiley & Sons, Inc., 9th Edition, 2013, pp: 56-70.
[2]Sachin Chaudhary et autres, **A Critical Review of Data Warehouse**, 5th IEEE International Conference on Advanced Computing & Communication Technologies [ICACCT-2011] ISBN 81-87885-03-3, p: 469.
[3] Henry C. Lucas, Jr., Op-cit, p : 249.
[4]Henry C. Lucas, Jr, **Information technology for management**, The Global Text Project is funded by the Jacobs Foundation, Zurich, Switzerland, 2009, p: 553.

- توقع تراجع العملاء باستخدام بيانات المبيعات للسنوات العشر الأخيرة؛
- إنشاء توقعات الطلب والمبيعات لتحديد المجالات التي يجب التركيز عليها في الربع القادم.

3- أسواق البيانات Data Marts

بينما يجمع مستودع البيانات قواعد البيانات عبر المؤسسة بأكملها، فإن متجر البيانات عادة ما يكون أصغر ويركز على موضوع أو قسم معين. متجر البيانات **data mart** هو مجموعة فرعية من مستودع البيانات، وعادة ما يتكون من منطقة موضوعية واحدة (مثل: التسويق، العمليات). يمكن أن يكون متجر البيانات إما تابعًا أو مستقلًا. متجر البيانات التابع هو مجموعة فرعية تُنشأ مباشرة من مستودع البيانات. وله مزايا استخدام نموذج بيانات متسق وتوفير بيانات ذات جودة عالية. تدعم متاجر البيانات التابعة مفهوم نموذج بيانات موحد على مستوى المؤسسة، ولكن يجب بناء مستودع البيانات أولاً. يضمن متجر البيانات التابع أن المستخدم النهائي يرى نفس إصدار البيانات الذي يصل إليه جميع مستخدمي مستودع البيانات الآخرين. يحد التكلفة العالية لمستودعات البيانات من استخدامها في الشركات الكبيرة. كبديل، تستخدم العديد من الشركات نسخة مصغرة منخفضة التكلفة من مستودع البيانات تُعرف باسم متجر البيانات المستقل. متجر البيانات المستقل هو مستودع صغير مصمم لوحدة أعمال استراتيجية أو قسم.

4- مستودعات البيانات المؤسسية Enterprise Data Warehouses (EDW)

مستودع البيانات المؤسسي (EDW) هو مستودع بيانات كبير يتم استخدامه عبر المؤسسة لدعم اتخاذ القرار. الطبيعة الواسعة النطاق توفر تكامل البيانات من العديد من المصادر في تنسيق قياسي لتطبيقات ذكاء الأعمال ودعم اتخاذ القرار الفعالة .تُستخدم مستودعات البيانات المؤسسية لتوفير البيانات للعديد من أنواع أنظمة دعم اتخاذ القرار، بما في ذلك إدارة علاقات العملاء(CRM) ، وإدارة سلسلة الإمداد(SCM) ، وإدارة أداء الأعمال(BPM) ، ورصد نشاط الأعمال(BAM) ، وإدارة دورة حياة المنتج (PLM)، وإدارة الإيرادات، وأحيانًا حتى أنظمة إدارة المعرفة(KMS) .

5- خصائص مستودعات البيانات

مستودعات البيانات (Data Warehouses) تُعد من الركائز الأساسية في نظم دعم القرار وتحليل البيانات في المؤسسات

- **التركيز على الموضوع**:(Subject-Oriented) : مستودع البيانات يُنظم البيانات حول مواضيع رئيسية تهم النشاط المؤسسي، مثل: العملاء، المنتجات، المبيعات، الأرباح، وليس وفقًا لعمليات يومية مثل إضافة طلب أو تحديث فاتورة، والهدف من هذه الخاصية هو تسهيل تحليل البيانات واتخاذ قرارات إستراتيجية بناءً على مواضيع محددة بدلاً من تتبع العمليات اليومية.

مثال: بدلاً من تتبع حركة كل منتج بشكل فردي، يُبنى المستودع ليُمكن من تحليل "أداء فئة منتجات" عبر الزمن.

– **الموضوعية**(Subject-Oriented) : مستودع البيانات يُنظم حول موضوعات رئيسية مثل: العملاء، المنتجات، المبيعات، وليس حول العمليات اليومية كالمعاملات الفردية.

– **التكامل**(**Integrated**) : يتم جمع البيانات من مصادر متعددة (مثل قواعد بيانات مختلفة، تطبيقات، ملفات)، وتوحيدها بصيغة موحدة داخل المستودع، حيث تشمل عملية التكامل: التوحيد في التسمية، وحدات القياس، الصيغ، وأنواع البيانات.

– **الثبات**(**Non-Volatile**) : البيانات في مستودع البيانات لا يتم تحديثها أو حذفها بانتظام كما في قواعد البيانات التشغيلية. تُضاف البيانات عبر عمليات تحميل مجدولة(ETL) ، وتُستخدم للتحليل وليس للتعديل المستمر.

– **البُعد الزمني(Time-Variant)** : البيانات تحتوي على أبعاد زمنية تُظهر التغيرات التاريخية، مما يتيح تتبع الاتجاهات وتحليل الأداء بمرور الوقت.

– **دعم اتخاذ القرار(Decision Support-Oriented)** : مصممة خصيصًا لتلبية احتياجات المستخدمين من التحليل، لا لإدارة العمليات اليومية، حيث تُستخدم في إعداد التقارير، ذكاء الأعمال(BI) ، واستخراج البيانات.(Data Mining)

– **الجودة العالية للبيانات(High Data Quality)** : بما أن الهدف الأساسي هو التحليل، يجب أن تكون البيانات نظيفة، دقيقة، وكاملة قدر الإمكان. يُستخدم عادة نظام ETL لضمان جودة البيانات أثناء نقلها من مصادرها إلى المستودع.

أمثلة على أدوات مستودعات البيانات:

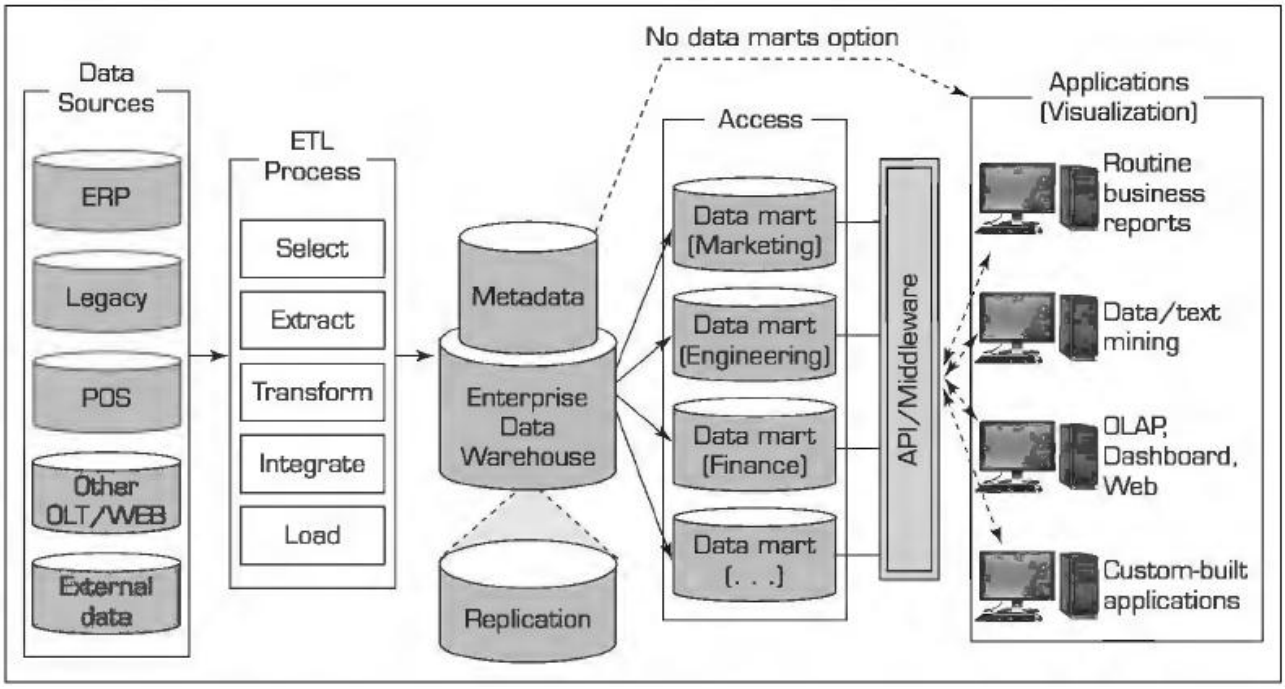
- Amazon Redshift
- Google BigQuery
- Snowflake
- Microsoft Azure Synapse Analytics
- Oracle Exadata

6- الشكل العام لمستودع البيانات

يمكن تصور مستودع البيانات على أنه يتكون من عدة مكونات رئيسية :

1. **المصدر** :حيث تأتي البيانات من أنظمة مختلفة مثل قواعد البيانات.
2. **عمليات** :**ETL** وهي العمليات التي تقوم باستخراج البيانات، وتحويلها إلى تنسيق مناسب، وتحميلها إلى مستودع البيانات.
3. **مستودع البيانات** :حيث يتم تخزين البيانات بشكل منظم، وغالبًا ما يتم استخدام نماذج مثل **Snowflake Schema** أو **Star Schema** لتسهيل الوصول إلى البيانات.
4. **أدوات التحليل والتقارير** :تستخدم لتحليل البيانات المستخرجة من مستودع البيانات.

الشكل رقم (01):إطار عمل مستودع البيانات



يتم بناء معظم مستودعات البيانات باستخدام أنظمة إدارة قواعد البيانات العلائقية (RDBMS) وهذه الأنظمة الأكثر استخداما:

Oracle (Oracle Corporation, **oracle.com**)

SQL Server (Microsoft Corporation, microsoft.com/sql)

DB2 (IBM Corporation, **http://www-Ol.ibm.com/software/data/ db2/**)

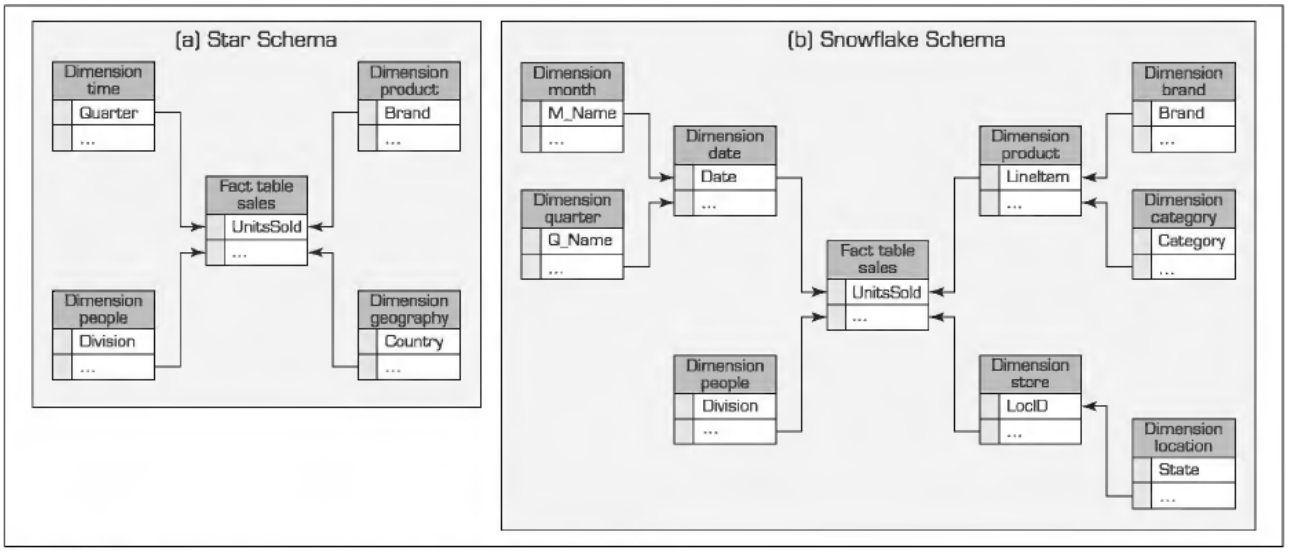
7- تمثيل البيانات في مستودع البيانات

يظهر هيكل مستودع البيانات النموذجي في الشكل رقم (01) العديد من التباينات الممكنة في بنية مستودع البيانات بغض النظر عن البنية، فإن تصميم تمثيل البيانات في مستودع البيانات يعتمد دائمًا على مفهوم النمذجة البُعدية. النمذجة البُعدية هي نظام يعتمد على الاسترجاع يدعم الوصول إلى الاستعلامات ذات الحجم الكبير. يجب تصميم تمثيل البيانات وتخزينها في مستودع البيانات بطريقة لا تقتصر فقط على استيعابها، بل تعزز أيضًا معالجة الاستعلامات المعقدة متعددة الأبعاد. غالبًا ما تكون مخطط النجمة Star Schema ومخطط رقاقة الثلج Snowflake Schema هما الوسيلتان اللتان يتم من خلالهما تنفيذ النمذجة البُعدية في مستودعات البيانات.

- **مخطط النجمة Star Schema**: هو الأكثر استخدامًا والأبسط في أسلوب النمذجة البُعدية. يحتوي مخطط النجمة على جدول حقائق مركزي محاط ومتصلاً بعدة جداول أبعاد يحتوي جدول الحقائق على السمات الوصفية اللازمة لإجراء تحليل القرار وإعداد تقارير الاستعلام، وتُستخدم المفاتيح الأجنبية للربط بجداول الأبعاد. تتكون سمات تحليل القرار من مقاييس الأداء، ومؤشرات التشغيل، والمقاييس المجمعة (مثل: أحجام المبيعات، ومعدلات الاحتفاظ بالعملاء، وهوامش الربح، وتكاليف الإنتاج، ومعدلات الفشل، وما إلى ذلك)، وجميع المقاييس الأخرى اللازمة لتحليل أداء المنظمة. بعبارة أخرى، يتناول جدول الحقائق بشكل أساسي ما يدعمه مستودع البيانات لتحليل القرار.

– **مخطط رقاقة الثلج Snowflake Schema** : هو ترتيب منطقي للجداول في قاعدة بيانات متعددة الأبعاد. يرتبط ارتباطًا وثيقًا بمخطط النجمة، يتم تمثيل مخطط الثلج بواسطة جداول حقائق مركزية (عادةً ما تكون واحدة فقط) متصلة بعدة أبعاد. في مخطط الثلج، تكون الأبعاد مُنظمة إلى جداول متعددة ذات صلة، بينما تكون أبعاد مخطط النجمة غير مُنظمة، حيث يتم تمثيل كل بعد بواسطة جدول واحد.

الشكل رقم (02):(a) The Star Schema, (b) the Snowflake Schema



8– مستودع البيانات مقابل قاعدة البيانات

الاختلافات الرئيسية بين مستودعات البيانات وقواعد البيانات:

الفرق الأكثر أهمية بين قواعد البيانات ومستودعات البيانات هو كيفية معالجة البيانات، حيث تستخدم قواعد البيانات معالجة المعاملات عبر الإنترنت (OLTP) (OnLine Transactional Processing) لحذف وإدراج واستبدال وتحديث أعداد كبيرة من المعاملات القصيرة عبر الإنترنت بسرعة. يستجيب هذا النوع من المعالجة على الفور لطلبات المستخدمين، وبالتالي يتم استخدامه لمعالجة العمليات اليومية لشركة ما في الوقت الفعلي. على سبيل المثال، إذا أراد المستخدم حجز غرفة فندق باستخدام نموذج الحجز عبر الإنترنت، يتم تنفيذ العملية باستخدام OLTP[1].

تستخدم مستودعات البيانات المعالجة التحليلية عبر الإنترنت (OLAP) (OnLine Analytical Processing) لتحليل مجلدات ضخمة من البيانات بسرعة. تمنح هذه العملية المحللين القدرة على النظر إلى البيانات من وجهات نظر مختلفة.

يوفر OLAP للمستخدمين الوسائل اللازمة لاستكشاف وتحليل كميات كبيرة من البيانات، بما في ذلك الحسابات المعقدة وعلاقاتهم وتقديم النتائج بشكل مرئي في وجهات نظر مختلفة. يستخدم OLAP طريقة عرض متعددة الأبعاد للبيانات المجمعة لتوفير وصول سريع إلى المعلومات الاستراتيجية لمزيد من التحليل.[2]

[1]Manole VELICANU, **Gheorghe MATEI, Database Vs Data Warehouse**, Revista Informatica Economică, nr. 3 (43)/2007, pp: 1-5.
[2]Forsman, S, OLAP Council White Paper, http://www.olapcouncil.org/research/whtpaply.htm

8-1- أنواع OLAP

هناك ثلاثة أنواع رئيسية:

-MOLAP: التخزين متعدد الأبعاد: يخزن البيانات في مكعبات مسبقة البناء. سريع جدًا ولكن مكلف.

-ROLAP: يعتمد على قواعد البيانات العلائقية: أكثر مرونة، ولكن أبطأ منMOLAP .

-HOLAP: مزيج من الاثنين: يخزن الملخصات في مكعبات وتفاصيل البيانات في قواعد البيانات.

8-2- الفرق بين مستودع البيانات وقاعدة البيانات

في مايلي أهم الفروقات:

الجدول رقم (01): الفرق بين مستودع البيانات وقاعدة البيانات

الخاصية	قاعدة البيانات	مستودع البيانات
طريقة المعالجة	معالجة المعاملات عبر الإنترنت (OLTP)	المعالجة التحليلية عبر الإنترنت (OLAP)
بنية البيانات	البيانات أكثر دقة ولكنها بطيئة في استرجاعها.	البيانات أقل دقة ولكن يتم استرجاعها بسرعة.
الجدول الزمني للبيانات	البيانات الحالية في الوقت الحقيقي لجزء واحد من العمل	البيانات التاريخية لجميع أجزاء العمل
تحليل البيانات	التحليل بطيء بسبب العدد الكبير من عمليات ربط الجداول المطلوبة والإطار الزمني الصغير للبيانات المتاحة.	التحليل سريع وسهل نظرًا للعدد الصغير من عمليات ربط الجداول المطلوبة والإطار الزمني الواسع للبيانات المتاحة.
المستخدمون المتزامنون	دعم الآلاف من المستخدمين المتزامنين. ومع ذلك، يمكن لمستخدم واحد فقط تعديل كل جزء من البيانات في المرة الواحدة.	عدد قليل من المستخدمين المتزامنين.
مدة التشغيل	وقت تشغيل بنسبة 99.99%	يتم تضمين وقت التوقف عن العمل لاستيعاب عمليات التحميل الدورية للبيانات الجديدة.
التخزين	يقتصر على مصدر بيانات واحد من وظيفة عمل معينة.	جميع مصادر البيانات من جميع وظائف العمل
نوع الاستعلام	استعلامات المعاملات البسيطة	استعلامات معقدة للتحليل المتعمق
الهدف	تم تصميمه للتسجيل	تم تصميمه للتحليل

التوفر	البيانات متاحة في الوقت الحقيقي	يتم تحديث البيانات من أنظمة المصدر عند الحاجة
نوع البيانات	البيانات المخزنة في قاعدة البيانات محدثة.	يتم تخزين البيانات الحالية والتاريخية في مستودع البيانات.

9- تطبيقات قاعدة البيانات

الجدول رقم (02):حالات استخدام قاعدة البيانات

القطاع	الاستخدام
الاستخدام المصرفي	في القطاع المصرفي للحصول على معلومات العميل، والأنشطة المتعلقة بالحسابات، والمدفوعات، والودائع، والقروض، وبطاقات الائتمان، إلخ.
شركات الطيران	تستخدمه شركات الطيران للحجوزات وجدولة المعلومات.
الجامعات	لتخزين معلومات الطلاب وتسجيلات الدورات والكليات والنتائج.
الاتصالات	تخزين سجلات المكالمات، والفواتير الشهرية، وصيانة الرصيد، وما إلى ذلك.
المالية	تخزين المعلومات المتعلقة بالمبيعات وشراء الأسهم والسندات.
المبيعات والإنتاج	تخزين تفاصيل العملاء والمنتجات والمبيعات.
التصنيع	يتم استخدامه لإدارة البيانات لسلسلة التوريد وتتبع إنتاج العناصر، وحالة المخزون.
إدارة الموارد البشرية	التفاصيل حول رواتب الموظف، والخصومات، وما إلى ذلك.

10- تطبيقات مستودع البيانات

الجدول رقم (03): حالات استخدام مستودع البيانات

القطاع	الاستخدام
الطيران	يتم استخدامه لعمليات إدارة نظام الطيران مثل تعيين الطاقم، وتحليلات الطريق، إلخ.
الرعاية الصحية	يستخدم في وضع الاستراتيجيات والتنبؤ بالنتائج، وإنشاء تقارير علاج المريض، وما إلى ذلك.
قطاع التأمين	تستخدم مستودعات البيانات على نطاق واسع لتحليل أنماط البيانات واتجاهات العملاء وتتبع حركات السوق بسرعة.
الاتصالات	تستخدم مستودع البيانات في العروض الترويجية للمنتج، وقرارات المبيعات واتخاذ قرارات التوزيع.