Cybrarians Journal
ISSN 1687-2215
دورية إلكترونية فصلية محكمة متخصصة في مجال المكتبات والمعلومات
العدد 18، مارس 2009

مستودعات البيانات

أحمد عبد الله
مركز جمعة الماجد
دبي، الإمارات العربية المتحدة
ahmed_nasser2000@hotmail.com

المستخلص

دراسة حول مستودعات البيانات، تتناول الدراسة تعريف مستودعات البيانات ثم تعرف أهم خصائصها، ثم تقدم الدراسة خطوات إنشاء مستودعات البيانات، ثم تناقش الدراسة تخزين البيانات في المستودعات وعرض البيانات في المستودعات، وأخيرا توضح الصعوبات التي تواجه مستودعات البيانات.

الاستشهاد المرجعي بالبحث

أحمد عبد الله. مستودعات البيانات .- cybrarians journal .- ع 18 (مارس 2009) . - تاريخ الاتاحة < اكتب هنا تاريخ اطلاعك على الصفحة > . - متاح في : <أكتب هنا رابط الصفحة الحالية>

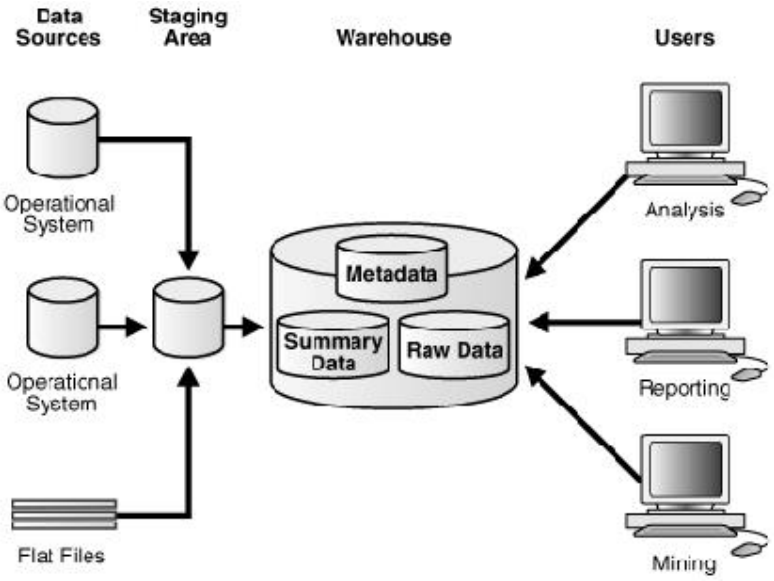
مقدمة

لقد ساعد الانتشار الواسع في مجال تكنولوجيا المعلومات إلى تضخم تلك المعلومات بصورة واسعة وكبيرة؛ حيث أصبحت عنصر مؤثرا وكبيراً في المجتمع بصفة عامة، لذلك كانت الضرورة الهامة والأساسية هي كيفية الاستفادة منها، ولكن مع الدخول والانتشار الواسع للإنترنت وتكنولوجيا الاتصالات أصبحت الوسيلة سهلة لعملية تبادل وانتشار المعلومات؛ حيث الحاجة إلى توثيق تلك المعلومات بطريقة آلية وتجميعها في قاعدة بيانات لإمكانية البحث في محتواها مع معالجة تلك البيانات باستخدام تقنيات وأدوات فعالية تمكننا من سرعة الوصول والإسترجاع

للمساعدة في الوصول إلى قرارات، من هنا ظهرت وتطورت قواعد البيانات تطوراً كبيراً منذ بدايتها حتي الآن، فمع بداية السبعينات ظهرت ما يعرف بقواعد البيانات الهرمية Hierarchique database وصمم هيكلها على أساس العلاقة بين السجلات التي تشكل هيكل شجري ومستويات هرمية ولهذا تعبر هذه التركيبة عن نمط العلاقات واحد – إلى كثير، وفي الثمانينات ظهرت قواعد البيانات العلائقية Relational database وهي من أكثر الأنماط استخداماً خصوصاً بعد ظهور وانتشار حزم نظم إدارة قواعد البيانات مع نظم الحاسب الشخصي، وفي التسعينات ظهرت قواعد البيانات الموجهة نحو الهدف Object-Oriented databases وهي الأكثر شيوعاً الآن لأنها تستطيع إدارة وسائط متعددة كما أنها تستخدم في تطبيقات شبكة الإنترنت؛ حيث أنها مفيدة في تخزين أنواع متعددة من البيانات، وهو ما يعرف بالجيل الرابع من قواعد البيانات، كل هذا أدي إلى بروز تحدي جديد في كيفية تحول البيانات من قواعد تخزين وبحث عن المعلومات إلى مخازن بيانات Data warehousing نستنتج منها المعرفة للمساعدة في اتخاذ القرار، لذلك كان لابد من معرفتها وخصائصها ووظائف تلك المستودعات وأهميتها في التنقب عن البيانات Data mining واستخراجها.

مفهوم مستودعات البيانات

عبارة عن "**مجموعة من بيانات دائمة تاريخية متكاملة تساعد في اتخاذ القرارات الإدارية**"[1]، للمساعد في الوصول إلى البيانات لأغراض التحليلات الزمنية واكتشاف المعرفة واتخاذ القرارات، فهي مصممة خصيصاً لإستخراج واستخلاص البيانات ومعالجتها وتقديمها وتمثيلها في صورة مناسبة لهذا الغرض، وتتضمن كميات ضخمة من البيانات تكون من مصادر مختلفة، أو من عدة قواعد مختلفة من أنظمة وأماكن مختلفة، كما يعرفها البعض الآخر "**على أنها نوع من قواعد البيانات تحوي كماً هائلاً من البيانات الموجهة للمساعدة على اتخاد القرارات داخل المؤسسة**"[2]، وتتميز تلك النوعية من قواعد البيانات بتطابق بنيتها الداخلية مع ما يحتاجة المستخدم من مؤشرات ومحاور التحليل، ويعرفها بندر الخيال[3] على أنها "**تلك النظم التي تهتم بدراسة وتحليل وعرض بيانات المنشأة بصورة تمكن الإدارة العليا من صنع القرار**"، ويتضح ذلك من معمارية بناء مستودع البيانات في شكل (1) الذي يوضح هيكلها العام.



شكل (1) نموذج مستودع البيانات

وقد برزت أهمية مستودعات البيانات مع وجود الشركات الكبرى؛ حيث أن كل قسم من تلك الشركات يدير قواعد بيانات خاصة به (تسويق، مالي، إدارة... الخ) والتي تتضمن كثيراً من البيانات المشتركة لذا تتطلب الحاجة إلى إيجاد بيانات متعلقة بموضوع محدد مثل العملاء على سبيل المثال من خلال البحث في قاعدة بيانات واحدة بدلاً من البحث في القواعد المختلفة، لذلك كان من الضروري تجميع البيانات في قاعدة واحدة تسمى مخزن البيانات مع الاحتفاظ بالقواعد الخاصة على أن تتم عملية تغذية دورية للمخازن إذا حدث أى تعديل أو تغير وتبديل في القواعد مع تصنيفها حسب الموضوع وذلك باستخدام برامج خاصة تسمى Middleware تستعمل بيانات دمج خاصة في كل موضوع تسمى Schema d integration يحدد فيها شكل وكيفية نقل البيانات.[4]

ويمكن القول بأن مستودعات البيانات هي عبارة عن قاعدة بيانات كبيرة وضخمة يمكن من خلالها الوصول إلى جميع المعلومات المتعلقة بمؤسسة أو شركة ما، ويودع في هذا المستودع البيانات الواردة من قواعد البيانات المختلفة كما يحتوي على العديد من قواعد البيانات وعلى معلومات من مصادر مختلفة، وتلك البيانات تكون في أشكال مختلفة ويمكن الوصول إليها من خلال الكمبيوتر الخادم، ويكون الوصول للمستودع غير مرئي بالنسبة للمستخدم الذي يستعمل أوامر بسيطة في التحليل واسترجاع وجمع البيانات، ويحتوى على بيانات تشير إلى كيفية تنظيمه،

وأماكن تلك المعلومات، وهي تستخدم لمساعدة الإدارة في كيفية اتخاذ القرارات داخل الشركات أو المؤسسات، كما تسمح للمؤسسة في تنظيم وتحديث وتنسيق تلك البيانات، وإظهار العلاقات بين المعلومات المجمعة من الإدارت المختلفة بالمؤسسة.

خصائص وصفات مستودعات البيانات

يمكن إجمال خصائصها في النقاط التالية:

1) تستخدم النموذج المتعدد الأبعاد Multidimensional model.
2) تدعم السلاسل الزمنية Time series وتحليل التوجهات Trends analysis اللذين يحتاجان لبيانات تاريخية لا تستطيع قواعد البيانات العادية أن توفرها.
3) تحديث البيانات وهو يتم كل فترة بواسطة أجزاء منه تختص بهذا الأمر.
4) استرجاع البيانات وتحليلها وهو صميم عملها وتهتم به أكثر.
5) مستويات تجميع Aggregation وأبعاد Dimensions غير محددة.
6) دعم معمارية العميل/الخادم وتعددية المستخدمين.
7) الاحتفاظ بكمية ضخمة من البيانات قد تصل إلى عدة تيرابايتات.
8) أنها محددة لموضوع Subject-Oriented على سبيل المثال في منشأة طبية ما تكون موضوعاتها تتعلق بالأطباء، المرضى، الممرضون، الأدوية... الخ فهي أشياء حسية ومعنوية.
9) أنها متكاملة Integrated من خلال وجود علاقة بين البيانات.
10) أنها غير قابلة للتعديل Non-Volatile أن تلك البيانات عند تحميلها في المستودعات تستخدم فقط للتحليل والدراسة والعرض.
11) أنها مرتبطة بعامل الوقت Time-Variant وهى من أهم نقطة في تلك البيانات.

خطوات بناء مستودعات البيانات

إن مشروع بناء مستودع للبيانات في مؤسسة ما ليس أمراً سهلاً، بل يتطلب فريق محترف في عملية تحليل الأنظمة والأعمال، مع وجود مهارات تقنية محترفة في ذلك الفريق، وهذه الخطوات كالتالي[5]:

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.

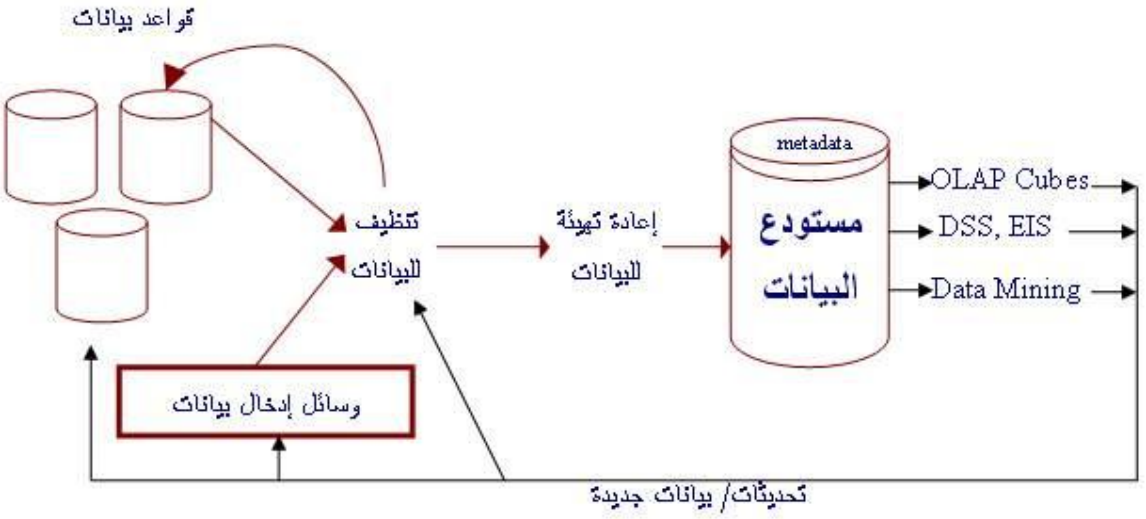
2. بناء مستودع البيانات Creating data warehouse وفيها ينشأ مستودع للبيانات والذي تحمل فيه بعد استخراجها وتنقيتها، وتصمم المستودعات دائما بحيث تسمح بوجود علاقات ذات أبعاد مختلفة.

3. تجزئة مستودع البيانات إلى مجموعة من متاجر البيانات Data marts 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 تلك المرحلة تحمل من Data preparation area إلى مستودع البيانات، ويتم فيها اختبار البيانات.

6. تحليل البيانات وإنشاء تطبيقات نظم دعم اتخاذ القرار OLAP & creating DSS applications وفي تلك 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، أيضاُ في تلك المرحلة تستخدم أدوات التنقب عن البيانات Data mining tools.



شكل (2) بناء مستودعات البيانات

هناك بعض القضايا التي يجب مراعاتها عند بناء مستودعات البيانات وهي كالتالي:

- 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.

معظم أو غالبية المؤسسات تعتمد بنية Three-tier architecture في عمل مستودعات البيانات، وفيها يقسم المستودع إلى ثلاثة مستويات على النحو الآتي:

- مستوى الأنظمة التشغيلية ومصادر البيانات المختلفة Data sources and operation systems.
- مستوع مستودع البيانات Data warehouses tier.
- مستوى تطبيقات العرض والتحليل OLAP systems.

وفي المستوى الأخير هناك الكثير من الشركات التي تقدم النظم والأدوات اللازمة لإنشاء مستودعات البيانات واستخلاص البيانات من مصادها، وأبرزها شركة أوركل وحلولها Oracle data mining, Oracle data warehouses builder.

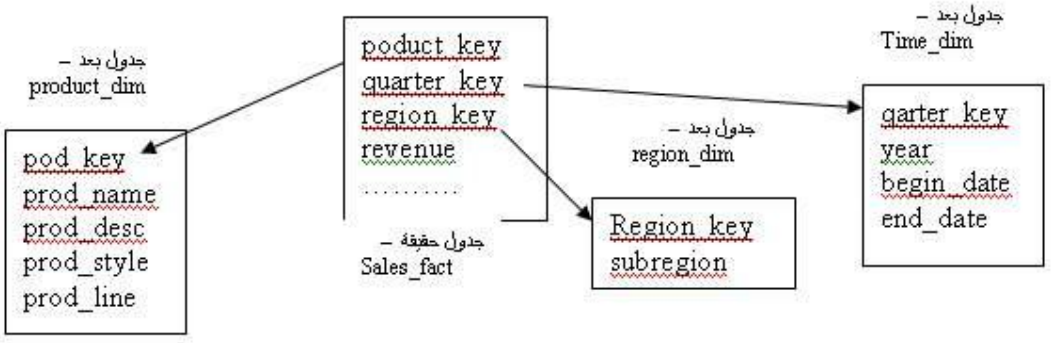
التخزين في مستودع البيانات

تخزين البيانات يتم بطرق متعددة الاتجاهات أو ما يعرف بالنموذج متعدد الأبعاد، وهي تستفيد من العلاقات بين البيانات لتسكينها في مصفوفات متعددة الأبعاد تسمى مكعبات البيانات Data cubes، وإذا استعملت قواعد البيانات العلائقية في مخازن البيانات وطريقة التحليل OLAP تسمى في تلك الحالة ROLAP relational، أو باستعمال قواعد بيانات متعددة الاتجاهات فتسمى Multidimensional OLAP، وعملية تخزين النموذج متعدد الأبعاد يتضمن نوعين من الجداول[6]:

- جدول البعد Dimension table وصفوفه تصف سمات البعد.
- جدول الحقيقة Fact table ويتضمن مقاييس أداء العمل، وتعرف بمؤشرات لجداول الأبعاد ويتضمن هذا الجدول البيانات.

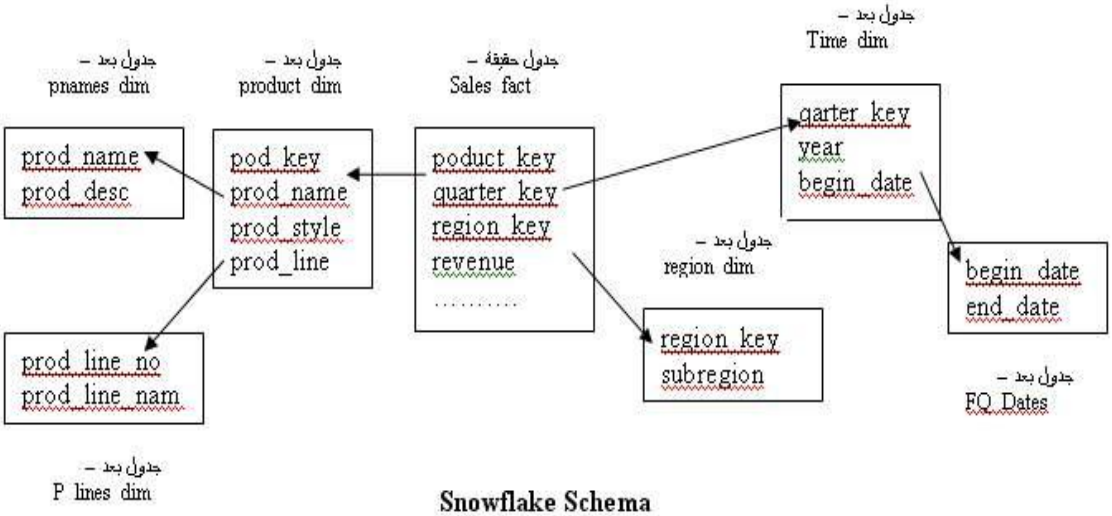
وتوجد بنيتين شائعتين لتعريف هذه الجداول، وهما:

أولاً: Star schema وتتكون من جدول الحقيقة مع جدول وحيد لكل بعد، ويعيبها احتياجها لمساحة كبيرة لتخزين البيانات.



شكل (3)

ثانياً: Snowflake schema وهى تعديل للجداول الأولى Star schema وذلك بتطبيق قوانين التطبيع Normalization عليها، ومشكلتها تضييعها لطاقة الجهاز المضيف في عمليات الربط.



شكل (4)

عرض النتائج في مستودعات البيانات

عملية العرض تستخدم طريقتين لتحليل تلك البيانات؛ **الطريقة الأولى**: وهي الحسابية حيث استعمال تحليل البيانات Analyze data من أجل إظهار النتائج بطريقة حسابية إحصائية (الجمع، الطرح، المعدل...الخ)؛ **الطريقة الثانية**: استخراج البيانات Data mining[*] وإظهار 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 لإتخاذ القرارات.

الصعوبات التي تواجه تطبيق مستودعات البيانات

- § الإنشاء يستغرق بعض الوقت ابتداءً من وضع الخطط حوله وحتى الانتهاء من تطبيقه.
- § إدارة المستودع صعبة نظرا للحجم الكبير وتعقيده، ويتطلب ذلك تدريباً للقائمين عليه وخصوصاً من ناحية مراعاة جودة البيانات.
- § التقدير لاحتياجات المستخدم قبل البدء في عملية الإنشاء.
- § ظهور أشكال ومواقع جديدة للبيانات بعد الإنتهاء من إنشاء المستودع يزيد من صعوبة إدارته.

الخلاصة

أصبح النمو الكبير والمتزايد للمعلومات في وقتنا الحاضر وخصوصا المنتشر بشكل كبير سواء على الإنترنت أو غيرها، يحتاج إلى وسيلة لتخزينه وتعديله وتنقيحه وتحديثه أول بأول للمحافظة على تلك المعلومات، لذلك ظهرت قواعد البيانات سواء المتخصصة أو العامة لنشر وإتاحة تلك البيانات بعدة طرق مختلفة، فأصبحت البيانات كثيرة فكان من الضروري إيجاد الطرق والسبل التي تكفل وتساعد الباحثين والمستفيدين للوصول السهل واليسير للحصول على ما يرغبون من بيانات، كل هذا يتطلب إنشاء مخازن أو مستودعات للبيانات للمساهمة ومساعدة الشركات سواء الكبرى أو الصغرى منها والتي أصبحت الآن تعتمد بشكل كبير على المعلومة والتي أصبحت سلعة لها أهميتها في الوقت الحاضر، ومن خلال توافر تلك المعلومات الضرورية والهامة بالطبع سوف تساعد الإدارة في أي مؤسسة في اتخاذ القرارات السريعة والهامة والدقيقة في نفس الوقت، كذلك المحافظة على الوقت الذي قد يضيع في سبل الوصول والحصول على البيانات إذا لم تكن مرتبة أو منظمة بطريقة تسهل الوصول إليها، من خلال تلك المستودعات يمكن استنباط الطرق التي قد تمكننا من التنقيب والبحث عن البيانات من أجل زيادة المعرفة.

المراجع

([1]) وليد محمد الحسن العمر، عبدالباسط عبدالصمد آدم. مستودعات البيانات. ص. 2.
http://www.walcom.jeeran.com/dw [Cited in: January 20, 2009]
([2]) http://ar.wikipedia.org [Cited in: January 18, 2009]
([3]) بندر الخيال. نبذة عن مستودعات البيانات في الحاسب الآلي. ص.2
http://www.elaph.com/Web/Technology/2008/10/373078.htm [Cited in: January 18, 2009]
([4]) محمود رمال. دور التقنيات الحديثة لقواعد المعلومات في بناء مجتمع المعلومات العربي
http://www.arabcin.net/arabic/5nadweh/pivot_3/new_technique.htm [Cited in: January 17, 2009]
([5]) بندر الخيال. مصدر سابق. ص. 2-3.
([6]) وليد محمد الحسن العمر، عبدالباسط عبدالصمد آدم. مصدر سابق. ص. 3-4.
([*]) هى عملية الوصول أو الحصول على المعرفة المهمة والجديدة من خلال كميات كبيرة من البيانات سواء حسب نماذج حسابية أو إحصائية أو منطقية.