

جامعة طرابلس
كلية تقنية المعلومات
قسم نظم المعلومات

المقرر الدراسي ITIS406

أنظمة دعم القرار

Decision support systems



المحاضرة الخامسة

اعداد: أ.فاطمة بشير القاضي
algadyfatma@gmail.com

مقدمة

إن الحاجة إلى تجميع البيانات من مصادرها المختلفة، وعدم قدرة أنظمة المعالجة الإجرائية transaction-processing systems على إنجاز تحليلات رفيعة المستوى للبيانات، جعل استخدام هذه الأنظمة غير فعال لدعم القرارات ولاسيما مع ازدياد حجم البيانات وتعقيدها.

تهدف أنظمة دعم القرار إلى تزويد مستخدميها بالمعلومات من أجل مساعدتهم في اتخاذ القرارات الصائبة، والقيام بالأعمال بشكل أكثر كفاءة، وهي تكنولوجيا جديدة ومتنامية في مجال حل المشكلات وتقييم الأداء وصنع القرار والتخطيط الاستراتيجي.

تابع مقدمة

ولحل مشكلة وفرة البيانات تستخدم أنظمة دعم القرار عدة تقنيات حديثة (ذكاء الاعمال) للمساعدة في عملية التحليل أهمها:

- مستودع البيانات Data warehousing
- تنقيب البيانات Data mining
- المعالجة التحليلية عبر الإنترنت Online analytical processing (OLAP)
- العرض المرئي للبيانات Data visualization

مستودع البيانات Data warehouse

تهدف تقنية مستودعات البيانات إلى التأكد من كون البيانات الملائمة متاحة للمستخدم المناسب في الوقت المناسب، وهي من أقوى أدوات دعم القرار، ظهرت فكرتها أوائل التسعينات وذلك كتحد لتحويل قواعد البيانات من قواعد تخزين وبحث عن المعلومة إلى مخازن تستنتج المعرفة وتساعد في اتخاذ القرار، وبالتالي نجد أن اعتماد نظم دعم القرار باستخدام تقنية مستودع البيانات يمكن من تتبع أثر البيانات وتزويدنا بالمعلومات المطلوبة بطريقة مباشرة وسريعة وذات معنى وهدف. تمكن المدراء من رؤية البيانات من وجهات نظر مختلفة مع تقليل كبير لزمان الاستعلام، ومن ثم صنع قرارات سريعة وتنافسية.

مستودعات البيانات: هي تلك النظم التي تهتم بدراسة بيانات الشركة وتحليلها وعرضها بصورة تمكن الإدارة العليا من صنع القرارات.

تابع مستودع البيانات

تمتاز البيانات المخزنة بمستودعات البيانات بالميزات الأربع الآتية:

1. موضوعية التوجه Subject-oriented:

إذ يتم تنظيم بيانات مستودع البيانات تبعاً لموضوع أو غرض أو شريحة معينة؛ مثل الزبائن، الموردون، المنتجات، المبيعات، وكلفة الصيانة والإصلاح، لذلك فإن مستودعات البيانات تعطي عرض بسيط وموجز وفي نفس الوقت ثري لموضوع معين، بعد استبعاد البيانات التي لا تعتبر جوهرية لعملية اتخاذ القرار، وتهتم بنمذجة البيانات modeling و تحليلها analysis من أجل صناع القرار من دون التركيز على الأعمال اليومية أو المعالجات الإجرائية اليومية في الشركة.

تابع مستودع البيانات

2. موحدة Integrated:

تبنى مستودعات البيانات من خلال تجميع ودمج البيانات من مصادر مختلفة وغير متجانسة، والتي تختلف أنواع ملفات الإلكترونية (قواعد البيانات العلائقية relational flat files، databases، وسجلات المعاملات المباشرة on-line transaction records، وغيرها)، لذلك فإن تطبيق تقنيات تنظيف البيانات وتوحيدها يضمن التأكد من التوافق والتناغم بين الأجزاء المختلفة.

مثل ضمان الاتساق في تسمية الصفات، نظام التشفير المستخدم، مقاييس وانواع الصفات، وغيرها من العمليات تتم على البيانات التي تم تجميعها من مصادر مختلفة.

تابع مستودع البيانات

3. التباين الزمني Time-variant:

من أجل الوصول إلى معرفة توجهات الأعمال يطلب المحللون كمية كبيرة من البيانات التاريخية للخمس أو العشر السنوات الماضية، لذلك فإن التركيز في مستودعات البيانات يكون على التغيرات التي تحدث عبر الزمن، وهو ما يقصد بالتباين الزمني، وهذا يجعل كل عنصر أو كائن في مستودع البيانات مرتبط بشكل ضمني أو صريح بعنصر الزمن، علي النقيض في قواعد البيانات التشغيلية قد ترتبط أو لا ترتبط عناصر البيانات فيها بعنصر الزمن.

تابع مستودع البيانات

4. مستقرة Nonvolatile:

إن البيانات غير قابلة للتعديل عند تحميلها في مستودعات البيانات فهي تستخدم فقط للتحليل والدراسة والعرض. وباعتبار أن مستودع البيانات مخزن للبيانات منفصل فيزيائياً عن قواعد البيانات التشغيلية في الشركة organization's operational databases، فهو لا يحتاج إلى معالجة المعاملات transaction processing، عمليات النسخ الاحتياطي الدورية، وآليات التحكم المتزامنة، وإنما يحتاج فقط إلى عمليتان هما تحميل البيانات والوصول إليها.

تابع مستودع البيانات

4. مستقرة Nonvolatile:

إن البيانات غير قابلة للتعديل عند تحميلها في مستودعات البيانات فهي تستخدم فقط للتحليل والدراسة والعرض. وباعتبار أن مستودع البيانات مخزن للبيانات منفصل فيزيائياً عن قواعد البيانات التشغيلية في الشركة organization's operational databases، فهو لا يحتاج إلى معالجة المعاملات transaction processing، عمليات النسخ الاحتياطي الدورية، وآليات التحكم المتزامنة، وإنما يحتاج فقط إلى عمليتان هما تحميل البيانات والوصول إليها.

العناصر الرئيسية لمستودع البيانات

Basic Elements of the Data Warehouse

1. أنظمة المصدر :Data Source Systems

وتتضمن قواعد البيانات التشغيلية operational databases و المصادر الخارجية external sources.

تابع العناصر الرئيسية لمستودع البيانات

2. منطقة تجميع البيانات Data staging area:

وهي منطقة تخزين متوسطة تقع بين أنظمة المصدر ومستودع البيانات، وهي عادة منطقة تخزين مؤقتة حيث تحذف منها البيانات بعد تحميلها إلى المستودع بنجاح. وتستخدم هذه المنطقة من أجل إجراء العديد من المعالجات الرئيسية مثل: أرشفة بيانات المصدر وإعدادها، واستخلاص البيانات، وتنظيفها، وتوحيدها integration, وإلغاء النسخ المتطابقة، والتحويل transformation والتحميل loading والفهرسة، بالإضافة إلى فحص سلامة النوعية والتحديث كل فترة زمنية يشار إلى هذه المعالجات عادة بـ ETL (Extraction, Transformation, Loading) وهي موجودة ضمن أدوات تُدعى Data Warehouse Back-End Tools and Utilities وتضم المهام الآتية:

تابع العناصر الرئيسية لمستودع البيانات

❖ **استخلاص البيانات Data extraction**: وهي الخطوة الأولى في عملية جلب البيانات إلى مستودع البيانات، وتعني قراءة البيانات وفهمها من مصادر متعددة، ثم نسخ الأجزاء اللازمة إلى منطقة تجميع البيانات من أجل مواصلة العمل لاحقاً. وتمثل خدمات الاستخلاص الجهد الأكبر في مستودع البيانات، إذ تُصرف حوالي 60 % من ساعات تطوير مستودع البيانات على عمليات الاستخلاص ويعدّ تحديد البيانات المطلوب استخلاصها وأنواع الفلترة المطبقة التحدي الأكبر، لذلك لا بد من فهم متطلبات عملية الاستخلاص من أجل تحديد أنواع الخدمات المطلوبة. وتتضمن هذه المرحلة أيضاً عملية تنظيف البيانات **Data cleaning** حيث يتم الكشف عن الأخطاء في البيانات وتصحيحها إذا كان أمكن، كالتعامل مع العناصر المفقودة، والإقلال من التشويش **noise** بتحديد القيم المتطرفة.

تابع العناصر الرئيسية لمستودع البيانات

❖ **تحويل البيانات Data transformation**: عندما تستخلص البيانات من

نظام المصدر يتم تطبيق سلسلة من الأفعال من أجل تحويل البيانات إلى صيغ صالحة للتقديم وذات قيمة. يضم التحويل الخطوات الآتية:

- تنقية الحقول المختارة من البيانات القديمة غير المفيدة لمستودع البيانات.
- الجمع بين مصادر البيانات عن طريق المطابقة بين القيم المفتاحية.
- إنشاء مفاتيح بديلة لكل سجل بعدي لتجنب الاعتماد على مفاتيح معرفة قديماً حيث تؤكد عملية توليد المفاتيح البديلة سلامة العلاقة بين الجداول البعدية وجداول الحقيقة.

- بناء التجمعات aggregates من أجل تعزيز أداء الاستعلام المشترك.
- تحويل نوع البيانات: وذلك من شكل إلى آخر (رقمي، حرفي، تاريخ .. الخ).

تابع العناصر الرئيسية لمستودع البيانات

❖ **التحميل Load:** تحتاج خدمات التحميل إلى دعم متطلباتها قبل التحميل وبعده كإعادة توليد الفهارس للجدول، بالإضافة إلى مراعاة خصوصية وبناء كل هدف عند القيام بعملية التحميل. وتتضمن هذه المرحلة أيضا عملية التحديث Refresh حيث يتم نقل التحديثات التي طرأت مع مرور الزمن من مصادر البيانات إلى المستودع.

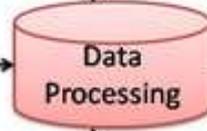
3. خادم العرض Presentation Server:

وهو المحرك الفيزيائي الذي ينظم بيانات مستودع البيانات ويخزنها، ويقدم إمكانية الاستعلام المباشر من قبل المستخدمين النهائيين End users وكاتبى التقارير وتطبيقات التحليل الأخرى.

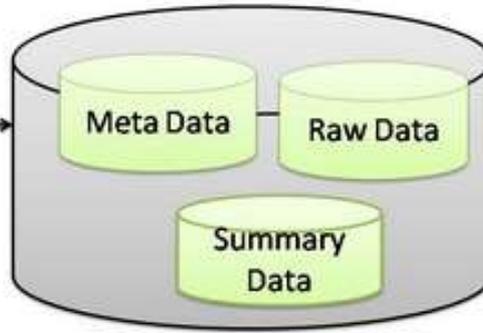
Data Sources



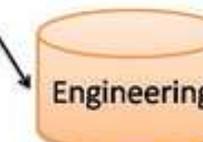
Data Staging
(ELT)



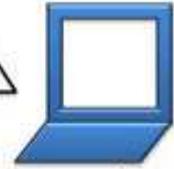
Data Warehouse



Data Marts



Users / Analysts



data warehouse architecture

بعض التعريفات الهامة

الملف المسطح Flat File: هو عبارة عن ملف لا يحتوي على أي روابط بملفات أخرى. قاعدة بيانات الملفات الثابتة هي قاعدة بيانات تخزن البيانات في ملف نصي عادي و تستخدم بشكل عام للقوائم المستقلة، لهذا عندما يجب أن تكون الملفات مرتبطة (العملاء بالأوامر ، البائعون بالمشتريات ، إلخ) ، يتم استخدام قاعدة البيانات العلائقية، وليس قاعدة بيانات الملفات الثابتة.

البيانات الوصفية Metadata: هي "بيانات / معلومات حول البيانات"، تساعدنا البيانات الوصفية على فهم بنية وطبيعة وسياق البيانات. البيانات الوصفية تسهل البحث عن البيانات واسترجاعها بسهولة، وتساعد أيضاً في الحفاظ على جودة وموثوقية البيانات.

تابع بعض التعريفات الهامة

البيانات الاولية Raw data: وتسمى ايضا البيانات الخام وهي البيانات التي يتم جمعها مباشرة من المصدر ولم تتم معالجتها أو تنظيمها أو تنظيفها أو تقديمها بشكل مرئي.

مجموعات البيانات Data Marts: هي مجموعة فرعية من مستودع البيانات تتعامل مع موضوع واحد فقط، وغالبًا ما يتم إنشاؤها وإدارتها بواسطة قسم أعمال واحد نظرًا. ولأنها موجهة نحو الموضوع **subject-oriented**، فإنها تأخذ عادةً بياناتها من عدد صغير من المصادر ، والتي يمكن أن تكون أنظمة تشغيل داخلية ، أو مستودع بيانات مركزيًا ، أو مصادر خارجية. عادةً ما تكون مكثفة وأقل تعقيدًا من مستودعات البيانات ، مما يسهل بنائها وصيانتها.

فوائد مستودع البيانات

Data warehouse benefits

- بإمكان المستخدم النهائي القيام بتحليل مكثف علي البيانات بطرق وتقنيات عدة.
- نسخة مركزة من البيانات المهمة للمؤسسة ومتاحة في اي وقت.
- مخازن البيانات تسمح بانسياب المعلومات من خوادم مكلفة (لتخزين قواعد البيانات التشغيلية) الي خوادم اقل تكلفة تتعامل مع مخزن جامع واحد، لذلك عمليات الاستفسار والتحليل تكون سهلة وسريعة.
- تحسين اداء الانظمة المعتمدة عليه، لان البيانات معالجة بالكامل.
- الوصول للبيانات يتم بطريقة سهلة لانها مجمعة في مكان واحد.
- مواكبة التغيرات المستمرة؛ عندما تظهر أسئلة جديدة أو تضاف بيانات جديدة فإن البيانات والتقنيات في المستودع لن تتغير أو تتعطل.

الفرق بين قواعد البيانات التشغيلية ومستودعات البيانات

1. الاستخدام: إن المهمة الرئيسة لأنظمة قواعد البيانات التشغيلية إنجاز المعالجة المباشرة للمعاملات والاستعلامات وتدعى هذه الأنظمة بأنظمة المعالجة الآنية للمعاملات (OLTP) on-line transaction processing وهي تغطي معظم العمليات اليومية في الشركة وتخزن عادة البيانات التابعة لعدة أسابيع أو أشهر لتقابل وبنجاح متطلبات المعاملات الحالية. أما أنظمة مستودعات البيانات فتدعم مستخدميها في عملية تحليل البيانات وصنع القرار، وتدعى هذه الأنظمة بأنظمة المعالجة التحليلية المباشرة (OLAP) on-line Analytical Processing وهي عبارة عن تطبيقات تجمع، وتدير، وتعالج، وتقدم البيانات ببناء متعدد الأبعاد من أجل ملاءمة الاحتياجات المتنوعة لمختلف المستخدمين، وهي تخزن عادة البيانات لعدة سنوات مما يدعم التحليل التاريخي Historical Analysis.

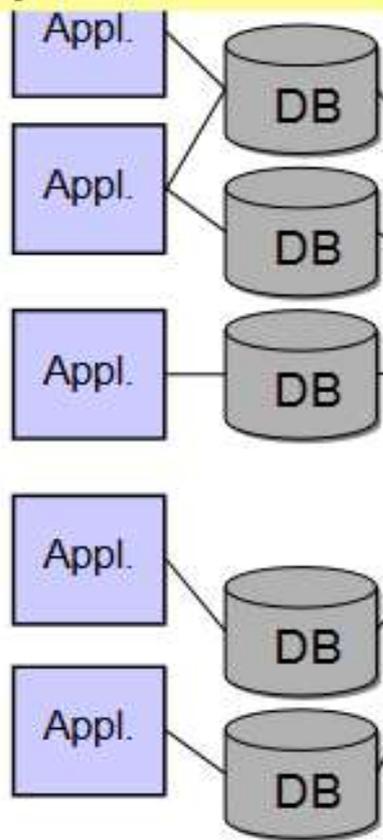
تابع الفرق بين قواعد البيانات التشغيلية ومستودعات البيانات

2. محتوى البيانات: تدير أنظمة OLTP البيانات الحالية والمفصلة، أما أنظمة OLAP فتدير كمية كبيرة من البيانات التاريخية، وتزودنا بتسهيلات من أجل تلخيص البيانات وتجميعها وتخزينها وإدارة المعلومات عند سويات مختلفة، مما يسهل استخدام البيانات الموجودة في المستودع من أجل صنع القرار. من جهة أخرى فإن أنظمة OLTP تمكن المستخدمين من إجراء تغييرات مباشرة على قاعدة البيانات، التي تكون Up to Date، على حين يتم تحديث مستودعات البيانات بشكل منتظم اعتماداً على تقنيات مؤتمتة في الاستخلاص والتحويل والتحميل، ولا يسمح للمستخدم النهائي أن يخزن البيانات مباشرة.

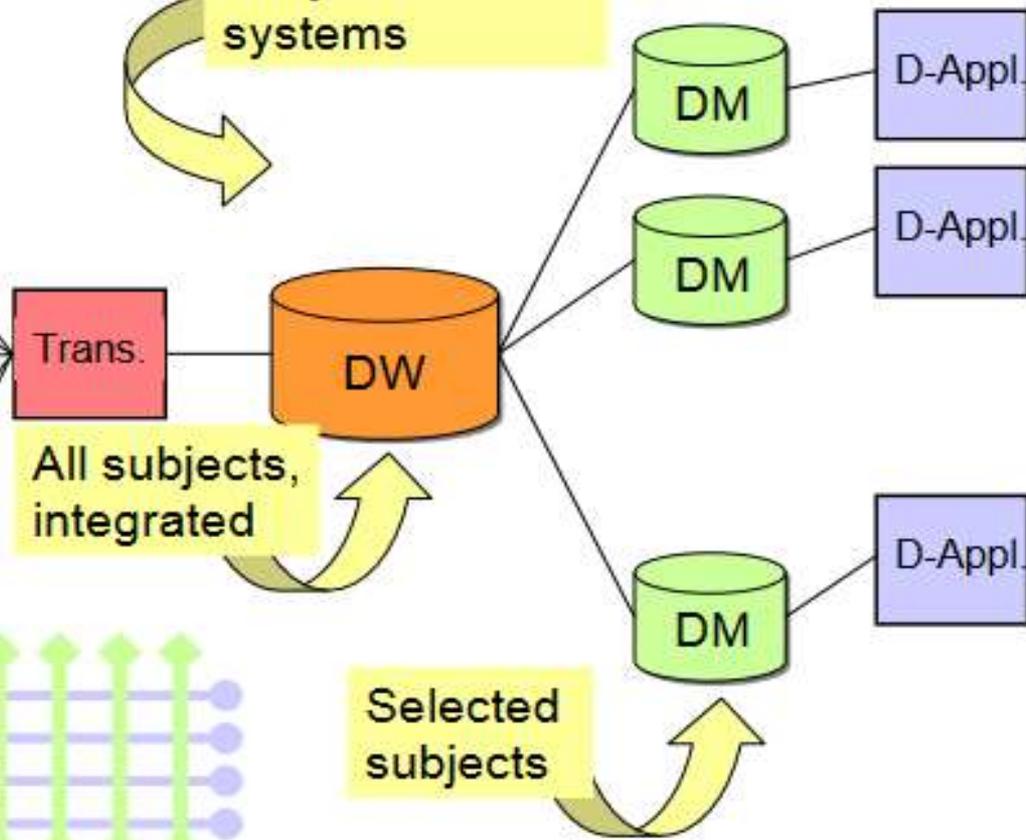
تابع الفرق بين قواعد البيانات التشغيلية ومستودعات البيانات

3. تصميم قاعدة البيانات: تبني أنظمة OLTP بشكل نموذجي وفق موديل العلاقات الكائنية (ER) Entity-Relationship Model ويكون تصميم قاعدة البيانات موجهاً نحو الوظائف Function-oriented، أما أنظمة OLAP فتبنى بشكل نموذجي وفق الموديل البعدي Multidimensional Data Model تصميم قاعدة البيانات موجهاً نحو القضايا والموضوعات الرئيسة Subject-Oriented وهذا يجعل مستودعات البيانات أسهل من حيث الاستعلام والفهم.

Function-oriented systems



Subject-oriented systems



All subjects, integrated

Selected subjects

