Data Warehousing

Main Data Warehousing Topics → المواضيع الرئيسية

- DW definition
- Characteristics of DW
- Data Marts
- ODS, EDW, Metadata
- DW Framework
- DW Architecture & ETL Process
- DW Development
- DW Issues

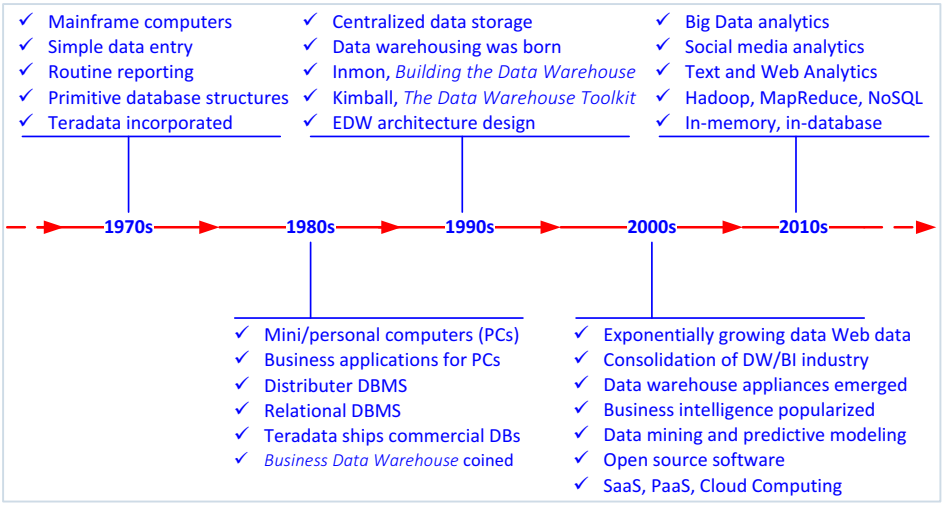
What is a Data Warehouse?

- A physical repository where relational data are specially organized to provide enterprise-wide, cleansed data in a standardized format
- "**The data warehouse** is a collection of integrated, subject-oriented databases designed to support DSS functions, where each unit of data is non-volatile and relevant to some moment in time"

ما هو مخزن البيانات ؟

- مخزن مادي او حقيقي حيث يتم تنظيم البيانات العلائقية خصيصًا لتوفير بيانات منقاه على مستوى المؤسسة في شكل قياسي.
- "**مخزن البيانات** عبارة عن مجموعة من قواعد البيانات المتكاملة والموجهة للموضوع المصممة لدعم وظائف DSS ، حيث تكون كل وحدة بيانات غير متطايرة وذات الصلة لبعض الوقت في الوقت المناسب ".

A Historical Perspective to Data Warehousing → منظور تاريخي لتخزين البيانات



1970s	✓ أجهزة الكمبيوتر المركزية ✓ إدخال بيانات بسيط ✓ التقارير الروتينية ✓ هياكل قاعدة البيانات البدائية ✓ توحيد تيراداتا
1980s	✓ اجهزة كمبيوتر صغيرة / شخصية (PCs) ✓ تطبيقات الأعمال لأجهزة الكمبيوتر ✓ موزع DBMS (نظام إدارة قواعد البيانات) (Database Management System) ✓ DBMSالعلائقية ✓ قاعدات البيانات تيراداتا للسفن التجارية ✓ صياغة مخزن بيانات الأعمال
1990s	✓ تخزين مركزي للبيانات ✓ ولد مخزن البيانات ✓ Inmon، بناء مخزن البيانات ✓ Kimball ، مجموعة أدوات مخزن البيانات ✓ تصميم هيكل (EDW (Enterprise Data Warehouse

2000s	✓ تزايد مستمر في بيانات الويب ✓ توطيد صناعة DW / BI ✓ ظهور أجهزة مخزن البيانات ✓ تعميم ذكاء الاعمال ✓ تعدين البيانات والنمذجة التنبؤية ✓ برنامج مفتوح المصدر ✓ SaaS، PaaS، الحوسبة السحابية (Software as a Service , Platform as a Service)
2010s	✓ تحليلات البيانات الكبيرة ✓ تحليلات وسائل الاعلام الاجتماعية ✓ تحليلات النص و الويب ✓ Hadoop، Mapreduce، NoSQL ✓ في الذاكرة ، في قاعدة البيانات

Characteristics of DWs (Data Warehouses)	خصائص DWs (مخازن البيانات)
❖ Subject oriented ❖ Integrated ❖ Time-variant (time series) ❖ Nonvolatile ❖ Summarized ❖ Not normalized ❖ Metadata ❖ Web based, relational/multi-dimensional ❖ Client/server, real-time/right-time/active...	❖ موجه الموضوع ❖ متكامل ❖ المتغير الزمني (السلسلة الزمنية) ❖ غير متطاير ❖ ملخص ❖ غير مطبع او غير منظم ❖ البيانات الوصفية ❖ مستندة على شبكة الإنترنت ، العلائقية / متعددة الأبعاد ❖ العميل / الخادم ، في الوقت الحقيقي / الوقت المناسب / نشط ...

Data Mart

A departmental small-scale "DW" that stores only limited/relevant data

- Dependent data mart
  - A subset that is created directly from a data warehouse
- Independent data mart
  - A small data warehouse designed for a strategic business unit or a department

سوق البيانات

صغير الحجم على مستوى"DW" يقوم بتخزين البيانات المحدودة / ذات الصلة فقط

- سوق البيانات المعتمدة
  - مجموعة فرعية يتم إنشاؤها مباشرة من مخزن البيانات
- سوق البيانات المستقلة
  - مستودع بيانات صغير مصمم لوحدة أعمال استراتيجية أو قسم

Other DW Components

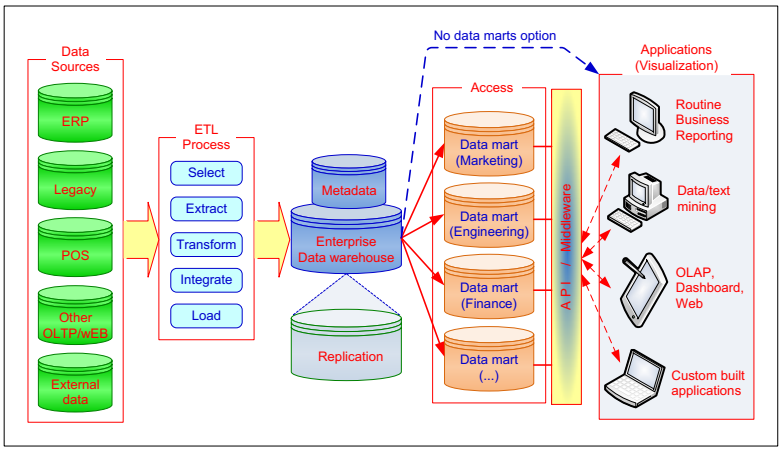
- **Operational data stores (ODS)**
  - A type of database often used as an interim area for a data warehouse
- **Oper marts** - an operational data mart.
- **Enterprise data warehouse (EDW)**
  - A data warehouse for the enterprise.
- **Metadata**: Data about data.
  - In a data warehouse, metadata describe the contents of a data warehouse and the manner of its acquisition and use

عناصر مخزن البيانات الأخرى

- مخازن البيانات التشغيلية (ODS)
  - نوع من قاعدة البيانات المستخدمة غالبًا كمنطقة مؤقتة لمخزن البيانات
- سجلات التشغيل - سوق البيانات التشغيلية.
- مخزن بيانات المؤسسة(EDW)

- مخزن البيانات للمؤسسة.
- البيانات الوصفية: بيانات حول البيانات.
  - في مخزن البيانات ، تصف البيانات الوصفية محتويات مخزن البيانات وطريقة اقتنائها واستخدامها

A Generic DW Framework → الهيكل العام لمخزن البيانات



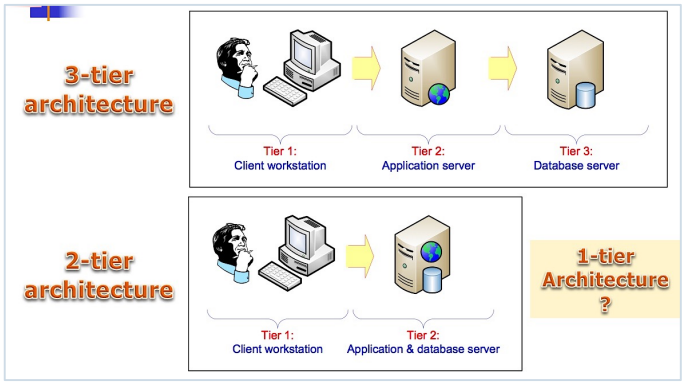
DW Architecture

- Three-tier architecture
  - Data acquisition software (back-end)
  - The data warehouse that contains the data & software
  - Client (front-end) software that allows users to access and analyze data from the warehouse
- Two-tier architecture
- First two tiers in three-tier architecture is combined into one
  - … sometimes there is only one tier?

بنية مخزن البيانات

- بنية من ثلاث طبقات
  - برنامج الحصول على البيانات (back-end)
  - مخزن البيانات الذي يحتوي على البيانات والبرامج
  - برنامج العميل (front-end) الذي يسمح للمستخدمين بالوصول إلى البيانات وتحليلها من المخزن

- بنية من طبقتان
- يتم دمج أول مستويين او طبقتين في بنية من ثلاث طبقات (three-tier architecture) و تكون في طبقة واحدة
  - ... في بعض الأحيان هناك طبقة واحدة فقط؟



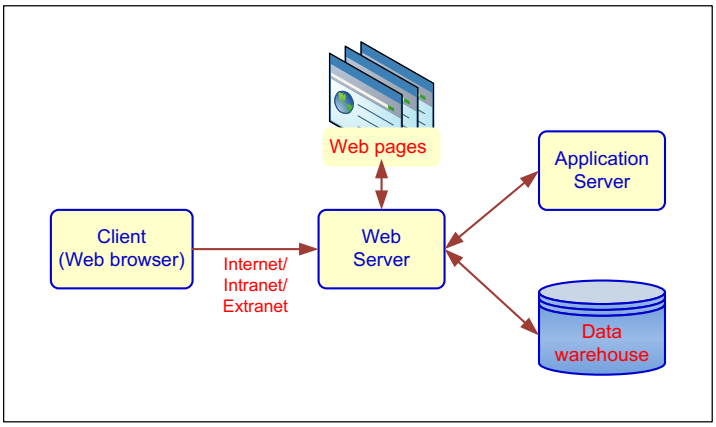
Data Warehousing Architectures

- Issues to consider when deciding which architecture to use:
  - Which database management system (DBMS) should be used?
  - Will parallel processing and/or partitioning be used?
  - Will data migration tools be used to load the data warehouse?
  - What tools will be used to support data retrieval and analysis?

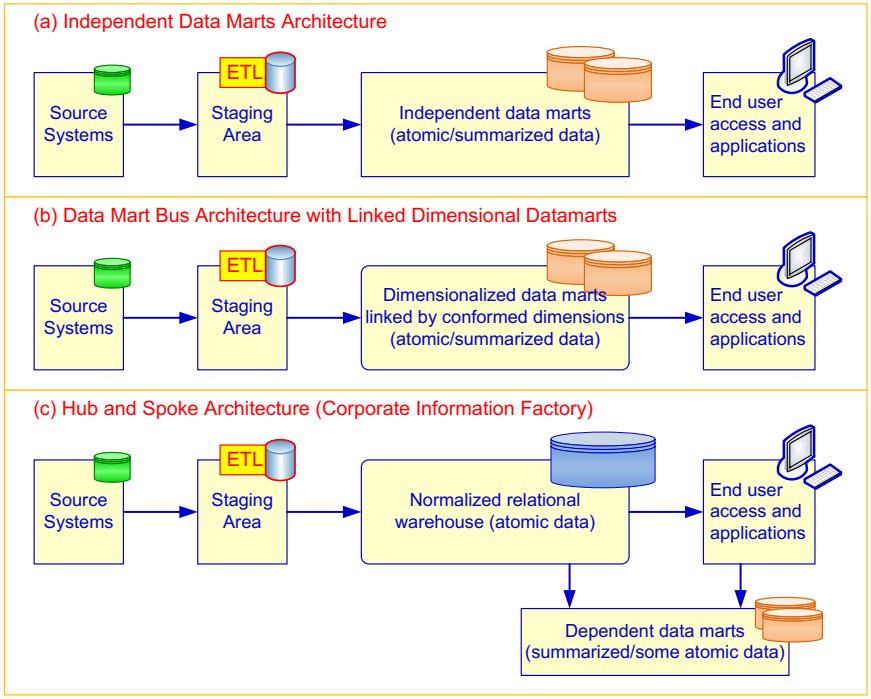
هياكل تخزين البيانات

- المشكلات التي يجب مراعاتها عند تحديد البنية المراد استخدامها:
  - ما هو نظام إدارة قواعد البيانات (DBMS) الذي ينبغي استخدامه؟
  - هل سيتم استخدام المعالجة المتوازية و/ أو التقسيم؟
  - هل سيتم استخدام أدوات تنقل البيانات لتحميل مخزن البيانات؟
  - ما الأدوات التي سيتم استخدامها لدعم استرجاع البيانات وتحليلها؟

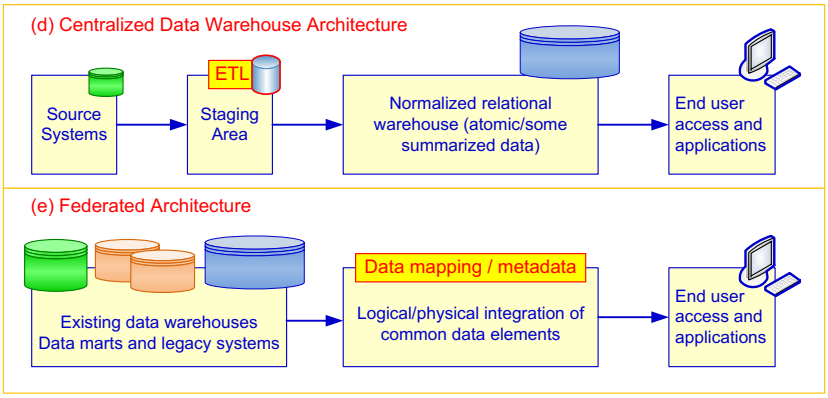
A Web-Based DW Architecture



Alternative DW Architectures



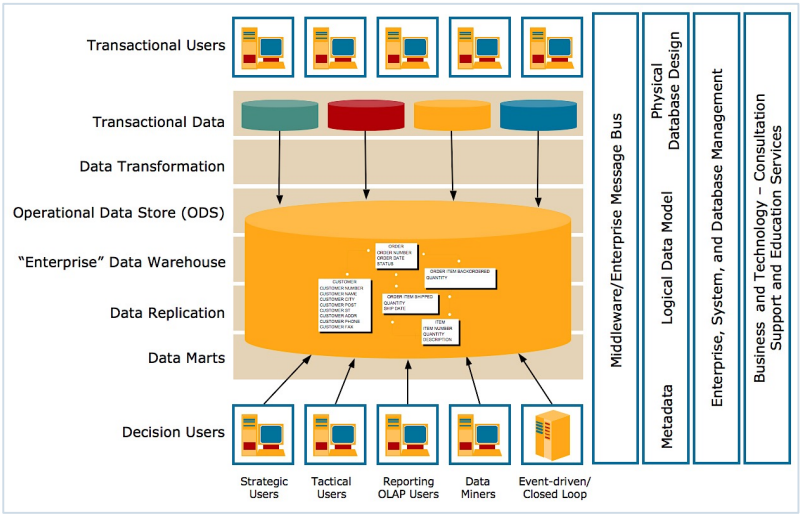
Alternative DW Architectures (cont.)



1. Each architecture has advantages and disadvantages! → كل بنية لها مزايا و عيوب
2. Which architecture is the best? → أي بنية هي الأفضل ؟

Ten factors that potentially affect the architecture selection decision	عشر عوامل من المحتمل انها تؤثر على قرار اختيار البنية
1. Information interdependence between organizational units	.1 ترابط المعلومات بين الوحدات التنظيمية
2. Upper management's information needs	.2 احتياجات الإدارة العليا للمعلومات
3. Urgency of need for a data warehouse	.3 الحاجة الملحة لمخزن البيانات
4. Nature of end-user tasks	.4 طبيعة مهام المستخدم النهائي
5. Constraints on resources	.5 القيود على الموارد
6. Strategic view of the data warehouse prior to implementation	.6 عرض استراتيجي لمخزن البيانات قبل التنفيذ
7. Compatibility with existing systems	.7 التوافق مع النظم القائمة
8. Perceived ability of the in-house IT staff	.8 القدرة المدركة لموظفي تكنولوجيا المعلومات الداخليين
9. Technical issues	.9 القضايا الفنية
10. Social/political factors	.10 العوامل الاجتماعية / السياسية

Teradata Corp. DW Architecture

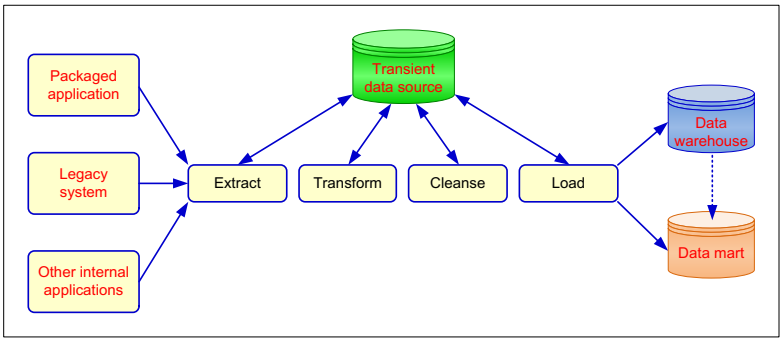


Data Integration and the Extraction, Transformation, and Load Process

- **ETL** = Extract Transform Load
- **Data integration**
  - Integration that comprises three major processes: data access, data federation, and change capture.
- **Enterprise application integration (EAI)**
  - A technology that provides a vehicle for pushing data from source systems into a data warehouse
- **Enterprise information integration (EII)**
  - An evolving tool space that promises real-time data integration from a variety of sources, such as relational or multidimensional databases, Web services, etc.

تكامل البيانات واستخراج وتحويل ومعالجة التحميل

- **ETL** = استخراج تحميل التحويلات
- **تكامل البيانات**
  - التكامل الذي يشتمل على ثلاث عمليات رئيسية: الوصول إلى البيانات ، واتحاد البيانات ، والتقاط التغيير.
- **تكامل تطبيقات المؤسسات (EAI)**
  - تقنية توفر وسيلة لدفع البيانات من أنظمة المصدر إلى مستودع او مخزن البيانات
- **تكامل معلومات المؤسسة (EII)**
  - مساحة أدوات متطورة تعد بدمج البيانات في الوقت الفعلي من مجموعة متنوعة من المصادر ، مثل قواعد البيانات العلائقية أو متعددة الأبعاد وخدمات الويب وما إلى ذلك.



ETL (Extract, Transform, Load)

- Issues affecting the purchase of an ETL tool
  - Data transformation tools are expensive
  - Data transformation tools may have a long learning curve
- Important criteria in selecting an ETL tool
  - Ability to read from and write to an unlimited number of data sources/architectures
  - Automatic capturing and delivery of metadata
  - A history of conforming to open standards
  - An easy-to-use interface for the developer and the functional user

استخراج ، تحميل ، التحويل

- المشكلات التي تؤثر على شراء أداة ETL
  - أدوات تحويل البيانات غالية الثمن
  - قد يكون لأدوات تحويل البيانات منحنى تعلم طويل
- معايير مهمة في اختيار أداة ETL
  - القدرة على القراءة والكتابة إلى عدد غير محدود من مصادر البيانات / البنيات
  - التقاط تلقائي للبيانات الوصفية وتسليمها
  - تاريخ من المطابقة للمعايير المفتوحة
  - واجهة سهلة الاستخدام للمطور والمستخدم الوظيفي

Data Warehouse Development

- Data warehouse development approaches
  - Inmon Model: EDW approach (top-down)
  - Kimball Model: Data mart approach (bottom-up)
  - Which model is best?
- Table 3.3 provides a comparative analysis between EDW and Data Mart approach
- One alternative is the hosted warehouse

تطوير مخزن البيانات

- مناهج تطوير مستودع او مخزن البيانات
  - نموذج Inmon: نهج EDW (من أعلى لأسفل)
  - نموذج Kimball: نهج مارت البيانات (من أسفل إلى أعلى)
  - أي نموذج هو الأفضل؟
- يقدم الجدول 3.3 تحليل مقارن بين EDW ونهج Data Mart
- بديل واحد هو مستضيف المخزن

Additional DW Considerations Hosted Data Warehouses	اعتبارات إضافية DW مستودعات البيانات المستضافة
Benefits: ○ Requires minimal investment in infrastructure ○ Frees up capacity on in-house systems ○ Frees up cash flow ○ Makes powerful solutions affordable ○ Enables solutions that provide for growth ○ Offers better quality equipment and software ○ Provides faster connections ○ … more in the book	الفوائد: ○ يتطلب الحد الأدنى للاستثمار في البنية التحتية ○ يحرر القدرة على الأنظمة الداخلية ○ يحرر التدفق النقدي ○ ينتج حلول قوية بأسعار معقولة ○ تمكن الحلول التي توفر للنمو ○ تقدم معدات وبرامج ذات جودة أفضل ○ يوفر اتصالات أسرع ○ ... أكثر في الكتاب

Representation of Data in DW

- **Dimensional Modeling**
  - A retrieval-based system that supports high-volume query access
- **Star schema**
  - The most commonly used and the simplest style of dimensional modeling
  - Contain a **fact table** surrounded by and connected to several **dimension tables**
- **Snowflakes schema**
  - An extension of star schema where the diagram resembles a snowflake in shape

تمثيل البيانات في مخزن البيانات

- **النمذجة البعدية**
  - نظام قائم على الاسترجاع يدعم وصول طلبات البحث ذات الحجم الكبير
- **مخطط النجوم**
  - النمط الأكثر استخداما وأبسط النمذجة الأبعاد
  - تحتوي على جدول حقائق محاط بطاولات الأبعاد ومتصلة بها
- **مخطط قطعة الثلج**
  - امتداد لمخطط النجوم حيث يشبه الرسم البياني لقطعة الثلج في الشكل

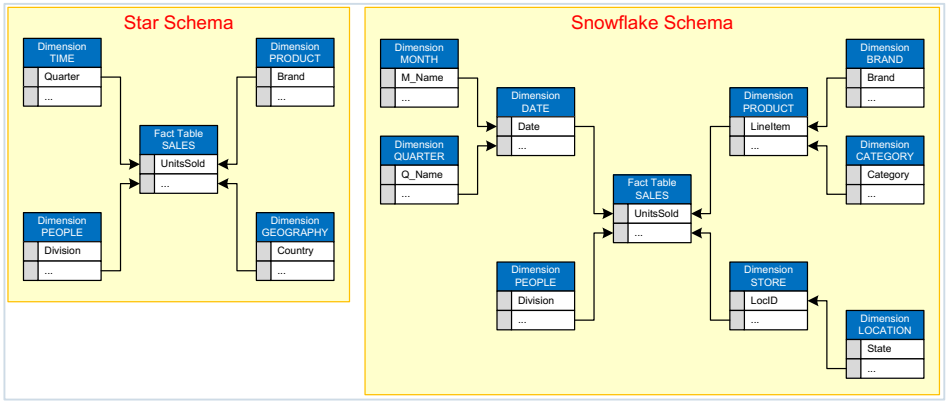
Multidimensionality

- The ability to organize, present, and analyze data by several dimensions, such as sales by region, by product, by salesperson, and by time (four dimensions)
- Multidimensional presentation
  - **Dimensions**: products, salespeople, market segments, business units, geographical locations, distribution channels, country, or industry
  - **Measures**: money, sales volume, head count, inventory profit, actual versus forecast
  - **Time**: daily, weekly, monthly, quarterly, or yearly

الابعاد المتعددة

- القدرة على تنظيم وعرض وتحليل البيانات بأبعاد متعددة ، مثل المبيعات حسب المنطقة ، حسب المنتج ، من قبل مندوب المبيعات ، وحسب الوقت (أربعة أبعاد)
- عرض متعدد الأبعاد
  - **الأبعاد**: المنتجات أو مندوبي المبيعات أو قطاعات السوق أو وحدات العمل أو المواقع الجغرافية أو قنوات التوزيع أو البلد أو الصناعة
  - **المقاييس**: المال ، وحجم المبيعات ، وعدد الرأس ، وجرد الربح ، الفعلي مقابل التوقعات
  - **الوقت**: يوميًا أو أسبوعيًا أو شهريًا أو ربع سنوي أو سنويًا

Star Vs Snowflake Schema



Analysis of Data in DW → تحليل البيانات في مخزن البيانات

- **OLTP vs. OLAP**…

○ **OLTP** (online transaction processing) • Capturing and storing data from ERP, CRM, POS, … • The main focus is on efficiency of routine tasks	○ **OLTP** (معالجة المعاملات عبر الإنترنت) • التقاط وتخزين البيانات من ERP، CRM، POS ، ... • ينصب التركيز الرئيسي على كفاءة المهام الروتينية
○ **OLAP** (Online analytical processing) • Converting data into information for decision support • Data cubes, drill-down / rollup, slice & dice, … • Requesting ad hoc reports • Conducting statistical and other analyses • Developing multimedia-based applications.	○ **OLAP** (معالجة تحليلية عبر الإنترنت) • تحويل البيانات إلى معلومات لدعم القرار • مكعبات البيانات ، والتنقيب / التراكم ، شريحة ونرد ، ... • طلب تقارير مخصصة • إجراء التحليلات الإحصائية وغيرها • تطوير التطبيقات القائمة على الوسائط المتعددة

OLTP vs. OLAP (cont.)

TABLE 3.5 A Comparison Between OLTP and OLAP

Criteria	OLTP	OLAP
Purpose	To carry out day-to-day business functions	To support decision making and provide answers to business and management queries
Data source	Transaction database (a normalized data repository primarily focused on efficiency and consistency)	Data warehouse or data mart (a nonnormalized data repository primarily focused on accuracy and completeness)
Reporting	Routine, periodic, narrowly focused reports	Ad hoc, multidimensional, broadly focused reports and queries
Resource requirements	Ordinary relational databases	Multiprocessor, large-capacity, specialized databases
Execution speed	Fast (recording of business transactions and routine reports)	Slow (resource intensive, complex, large-scale queries)

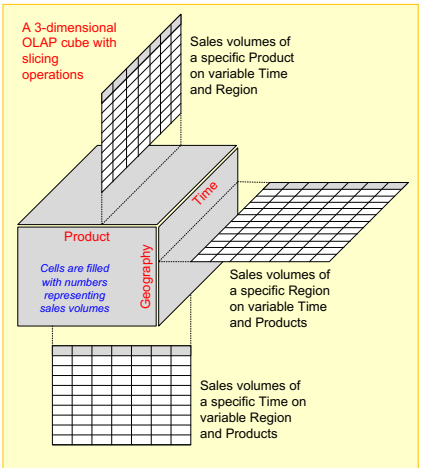
المعايير	OLTP	OLAP
الغرض	للقيام بمهام العمل اليومية	لدعم صنع القرار وتقديم إجابات لاستفسارات الأعمال والإدارة
مصدر البيانات	قاعدة بيانات المعاملات (مخزن البيانات المعتاد يركز في المقام الأول على الكفاءة والاتساق)	مخزن البيانات أو سوق البيانات (يركز مخزن البيانات غير العادي على الدقة والاكتمال)
التقارير	تقارير روتينية ودورية ومركزة بشكل دقيق	تقارير واستعلامات مخصصة ومتعددة الأبعاد ومركزة على نطاق واسع
متطلبات المصدر	قاعدة بيانات علائقية مألوفة	قاعدة بيانات متعددة المعالجات ، ذات سعة كبيرة ، متخصصة
سرعة التنفيذ	بسرعة (تسجيل المعاملات التجارية والتقارير الروتينية)	بطيء (استعلامات مكثفة ، معقدة ، الموارد مكثفة)

OLAP Operations

- **Slice** - a subset of a multidimensional array
- **Dice** - a slice on more than two dimensions
- **Drill Down/Up** - navigating among levels of data ranging from the most summarized (up) to the most detailed (down)
- **Roll Up** - computing all of the data relationships for one or more dimensions
- **Pivot** - used to change the dimensional orientation of a report or an ad hoc query-page display

عمليات OLAP

- شريحة - مجموعة فرعية من صفوفة متعددة الأبعاد
- النرد - شريحة على أكثر من بعدين
- الحفر لأسفل / لأعلى - التنقل بين مستويات البيانات التي تتراوح من أكثر تلخيص (أعلى) إلى الأكثر تفصيلاً (أسفل)
- لف او طي - حساب جميع علاقات البيانات لبُعد واحد أو أكثر
- Pivot - يُستخدم لتغيير اتجاه الأبعاد لتقرير أو عرض صفحة استعلام مخصصة



Slicing Operations on a Simple Tree-Dimensional Data Cube

Variations of OLAP

- **Multidimensional OLAP (MOLAP)**
  - OLAP implemented via a specialized multidimensional database (or data store) that summarizes transactions into multidimensional views ahead of time
- **Relational OLAP (ROLAP)**
  - The implementation of an OLAP database on top of an existing relational database
- Database OLAP and Web OLAP (DOLAP and WOLAP); Desktop OLAP,…

اختلافات OLAP

- **OLAP متعدد الأبعاد (MOLAP)**
  - تنفيذ OLAP عبر قاعدة بيانات متعددة الأبعاد متخصصة (أو مخزن بيانات) تلخص المعاملات في وجهات نظر متعددة الأبعاد في وقت مبكر
- **OLAP العلائقية (ROLAP)**
  - تنفيذ قاعدة بيانات OLAP أعلى قاعدة بيانات علائقية موجودة
- قاعدة بيانات OLAP و شبكة OLAP (WOLAP وDOLAP) و سطح المكتب OLAP

DW Implementation Issues	قضايا التنفيذ DW (مخزن البيانات)
Identification of data sources and governance	تحديد مصادر البيانات والحكم
Data quality planning, data model design	تخطيط جودة البيانات ، تصميم نموذج البيانات
ETL tool selection	اختيار أداة ETL
Establishment of service-level agreements	إنشاء اتفاقات مستوى الخدمة
Data transport, data conversion	نقل البيانات ، تحويل البيانات
Reconciliation process	عملية المصالحة
End-user support	دعم المستخدم النهائي
Political issues	مشاكل سياسية

Successful DW Implementation -- Things to Avoid

- Starting with the wrong sponsorship chain
- Setting expectations that you cannot meet
- Engaging in politically naive behavior
- Loading the data warehouse with information just because it is available
- Believing that data warehousing database design is the same as transactional database design
- Choosing a data warehouse manager who is technology oriented rather than user oriented

نجاح تنفيذ DW -- أشياء يجب تجنبها

- بدءاً من سلسلة الرعاية الخاطئة
- تحديد التوقعات التي لا يمكنك الوفاء بها
- الانخراط في سلوك سياسي ساذج
- تحميل مخزن البيانات بالمعلومات فقط لأنه متاح
- الاعتقاد بأن تصميم قاعدة بيانات لتخزين البيانات هو نفس تصميم قاعدة بيانات المعاملات
- اختيار مدير مستودع البيانات الذي يكون موجهًا للتكنولوجيا بدلاً من المستخدم

Failure Factors in DW Projects	عوامل الفشل في مشاريع DW
❖ Lack of executive sponsorship ❖ Unclear business objectives ❖ Cultural issues being ignored ○ Change management ❖ Unrealistic expectations ❖ Inappropriate architecture ❖ Low data quality / missing information ❖ Loading data just because it is available	❖ عدم وجود رعاية تنفيذية ❖ أهداف تجارية واضحة ❖ القضايا الثقافية يتم تجاهلها ○ تغيير الإدارة ❖ توقعات غير واقعية ❖ بنية غير لائقة ❖ جودة منخفضة للبيانات / المعلومات المفقودة ❖ تحميل البيانات لمجرد توافرها

Massive DW and Scalability

- Scalability
  - The main issues pertaining to scalability:
    - The amount of data in the warehouse
    - How quickly the warehouse is expected to grow
    - The number of concurrent users
    - The complexity of user queries
  - Good scalability means that queries and other data-access functions will grow linearly with the size of the warehouse

ضخامة DW وقابلية التطوير

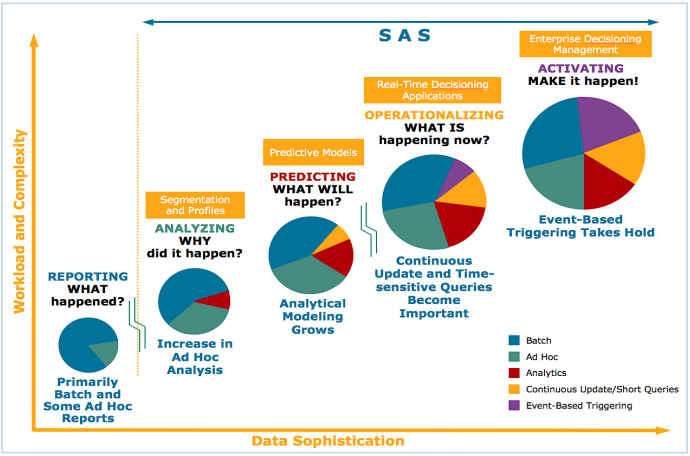
- قابلية التوسع
  - القضايا الرئيسية المتعلقة بقابلية التوسع:
    - كمية البيانات في المخزن
    - ما مدى سرعة نمو المخزن
    - عدد المستخدمين المتزامنين
    - تعقيد استعلامات المستخدم
  - تعني قابلية التوسع الجيد أن الاستعلامات ووظائف الوصول إلى البيانات الأخرى سوف تنمو بشكل خطي مع حجم المخزن

Real-Time/Active DW/BI

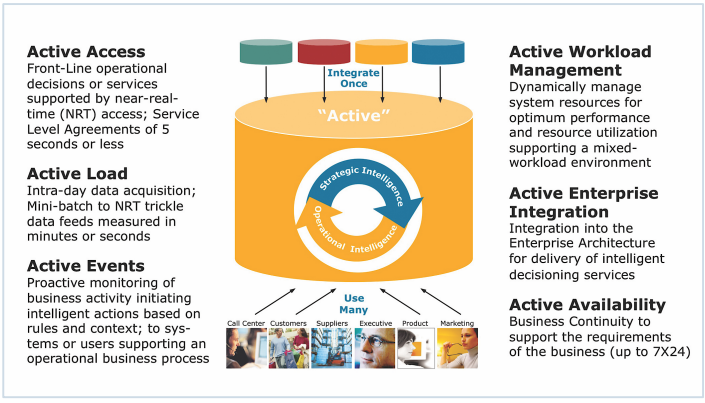
- Enabling real-time data updates for real-time analysis and real-time decision making is growing rapidly
  - Push vs. Pull (of data)
- Concerns about real-time BI
  - Not all data should be updated continuously
  - Mismatch of reports generated minutes apart
  - May be cost prohibitive
  - May also be infeasible

- ينمو بسرعة تحديث البيانات في الوقت الحقيقي لتحليل في الوقت الحقيقي وصنع القرار في الوقت الحقيقي
  - دفع Vs سحب (من البيانات)
- مخاوف بشأن BI في الوقت الحقيقي
  - لا ينبغي تحديث جميع البيانات بشكل مستمر
  - أدى عدم توافق التقارير إلى إنشاء دقائق منفصلة
  - قد تكون تكلفة باهظة
  - قد يكون أيضا غير قابل للتطبيق

Enterprise Decision Evolution and Data Warehousing



Real-Time/Active DW at Teradata



Traditional versus Active DW

Traditional Data Warehouse Environment	Active Data Warehouse Environment
Strategic decisions only	Strategic and tactical decisions
Results sometimes hard to measure	Results measured with operations
Daily, weekly, monthly data currency acceptable; summaries often appropriate	Only comprehensive detailed data available within minutes is acceptable
Moderate user concurrency	High number (1,000 or more) of users accessing and querying the system simultaneously
Highly restrictive reporting used to confirm or check existing processes and patterns; often uses predeveloped summary tables or data marts	Flexible ad hoc reporting, as well as machine-assisted modeling (e.g., data mining) to discover new hypotheses and relationships
Power users, knowledge workers, internal users	Operational staffs, call centers, external users

بيئة مخزن البيانات النشطة	بيئة مخزن البيانات التقليدية
✓ القرارات الاستراتيجية والتكتيكية ✓ النتيجة تقاس بالعمليات ✓ فقط البيانات التفصيلية الشاملة المتاحة في غضون دقائق هي المقبولة ✓ عدد مرتفع (1000 أو أكثر) من المستخدمين الذين يصلون إلى النظام ويستعلمون عنه في وقت واحد ✓ إعداد تقارير مخصصة مرنة ، بالإضافة إلى النمذجة المدعومة آليًا (مثل استخراج البيانات) لاكتشاف فرضيات وعلاقات جديدة ✓ الموظفين التشغيليين ومراكز الاتصال والمستخدمين الخارجيين.	✓ القرارات الاستراتيجية فقط ✓ النتيجة صعبة في بعض الأحيان للقياس ✓ عملات مقبولة يومية وأسبوعية وشهرية للبيانات. الملخصات في كثير من الأحيان مناسبة ✓ تزامن المستخدم المعتدل ✓ استخدام التقارير المقيدة للغاية لتأكيد أو التحقق من العمليات والأنماط الحالية ؛ غالبًا ما يستخدم الجداول الموجزة التي تم تطويرها مسبقًا أو سجلات البيانات ✓ مستخدمو الطاقة ، عمال المعرفة ، المستخدمون الداخليون

DW Administration and Security

- Data warehouse administrator (DWA)
  - DWA should…
    - have the knowledge of high-performance software, hardware and networking technologies
    - possess solid business knowledge and insight
    - be familiar with the decision-making processes so as to suitably design/maintain the data warehouse structure
    - possess excellent communications skills
- Security and privacy is a pressing issue in DW
  - Safeguarding the most valuable assets
  - Government regulations (HIPAA, etc.)
  - Must be explicitly planned and executed

إدارة DW والأمن

- مدير مستودع البيانات (DWA)
  - يجب ... DWA
    - الحصول على معرفة بالبرمجيات عالية الأداء ، وتقنيات الأجهزة والشبكات
    - تمتلك معرفة تجارية راسخة وفطينة
    - أن يكونوا على دراية بعمليات صنع القرار حتى يتم تصميم / صيانة هيكل مستودع البيانات بشكل مناسب
    - تمتلك مهارات اتصالات ممتازة
- الأمان والخصوصية مشكلة ملحة في DW
  - حماية الأصول الأكثر قيمة
  - اللوائح الحكومية (HIPAA ، إلخ)
  - يجب أن يتم تخطيطها وتنفيذها بشكل صريح

The Future of DW

- Sourcing…
  - *Web, social media, and Big Data*
  - Open source software
  - SaaS (software as a service)
  - Cloud computing
- Infrastructure…
  - Columnar
  - Real-time DW
  - Data warehouse appliances
  - Data management practices/technologies
  - In-database & In-memory processing New DBMS
  - Advanced analytics

مستقبل مخزن البيانات

- مصادر ...
  - الويب والوسائط الاجتماعية والبيانات الضخمة
  - برنامج مفتوح المصدر
  - SaaS (برنامج كخدمة)
  - حوسبة سحابية
- بنية تحتية...
  - عمودي
  - أجهزة مستودع البيانات
  - في قاعدة البيانات والمعالجة في الذاكرة نظام إدارة قواعد البيانات جديد
  - في الوقت الحقيقي DW
  - ممارسات / تقنيات إدارة البيانات
  - تحليلات متقدمة