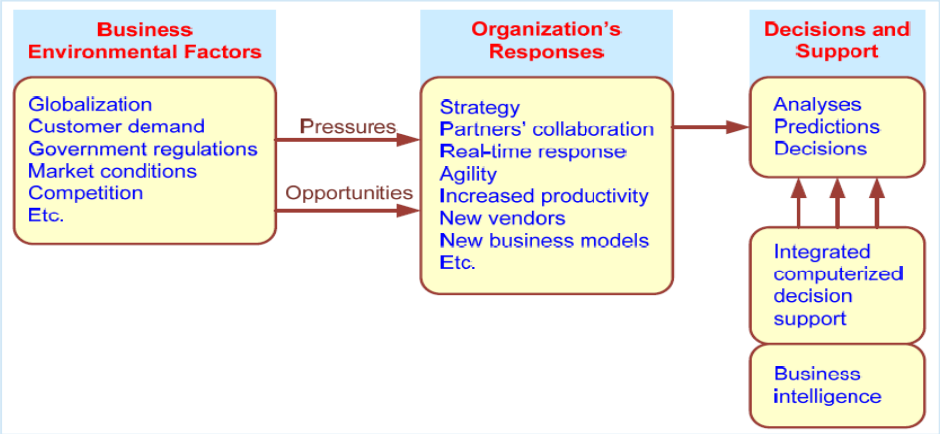
An Overview of Business Intelligence, Analytics, and Decision Support

Changing Business Environment & Computerized Decision Support

- Companies are moving aggressively to computerized support of their operations (Business Intelligence).
- Business Pressures–Responses–Support Model
  - **Business pressures** result of today's competitive business climate
  - **Responses** to counter the pressures
  - **Support** to better facilitate the process

تغيير بيئة العمل ودعم اتخاذ القرار بالكمبيوتر

- تتجه الشركات بقوة إلى الدعم المحوسب لعملياتها (ذكاء الأعمال).
- الضغوط التجارية - الاستجابة - نموذج الدعم
  - **الضغوط التجارية:** نتيجة للمناخ او البيئة التجارية التنافسية لليوم
  - **الاستجابة:** تكون على مواجهة الضغوط
  - **الدعم:** يكون على تسهيل وتيسير العملية بشكل أفضل.



The Business Environment

- The environment in which organizations operate today is becoming more and more complex, creating
  - Opportunities
  - Problems
    - Example: globalization.
- Business environment factors:
  - markets, consumer demands, technology, and societal…

بيئة الاعمال

- إن البيئة التي تعمل فيها المنظمات اليوم أصبحت أكثر تعقيدًا
  - الفرص
  - المشاكل
    - مثال عليهم : العولمة
- عوامل بيئة الاعمال:
  - الأسواق ومطالب المستهلكين والتكنولوجيا والمجتمع.

Business Environment Factors

FACTOR DESCRIPTION	وصف العوامل
❖ **Markets** ○ Strong competition ○ Expanding global markets ○ Blooming electronic markets on the Internet ○ Innovative marketing methods ○ Opportunities for outsourcing with IT support ○ Need for real-time, on-demand transactions	❖ **الأسواق** ○ تنافس قوي ○ توسيع الأسواق العالمية ○ تتفتح الأسواق الإلكترونية على الإنترنت ○ طرق تسويق مبتكرة ○ فرص الاستعانة بمصادر خارجية مع دعم تكنولوجيا المعلومات ○ الحاجة إلى معاملات في الوقت الحقيقي، حسب الطلب
❖ **Consumer demands** ○ Desire for customization ○ Desire for quality, diversity of products, and speed of delivery ○ Customers getting powerful and less loyal	❖ **مطالب المستهلكين** ○ الرغبة في التخصيص او بما يسمى الإنتاج حسب طلب العميل. ○ الرغبة في الجودة وتنوع المنتجات وسرعة التسليم ○ العملاء يملكون قوة وولاء اقل.
❖ **Technology** ○ More innovations, new products, and new services ○ Increasing obsolescence rate ○ Increasing information overload ○ Social networking, Web 2.0 and beyond	❖ **التكنولوجيا** ○ مزيد من الابتكارات والمنتجات الجديدة والخدمات الجديدة ○ الزيادة في معدل التقادم ○ الزيادة في المعلومات الزائدة ○ الشبكات الاجتماعية ، الويب 2.0 وما بعدها
❖ **Societal** ○ Growing government regulations and deregulation ○ Workforce more diversified, older, and composed of more women ○ Prime concerns of homeland security and terrorist attacks ○ Necessity of Sarbanes-Oxley Act and other reporting-related legislation ○ Increasing social responsibility of companies ○ Greater emphasis on sustainability	❖ **المجتمع** ○ تزايد اللوائح الحكومية وإلغاء القيود التنظيمية ○ القوى العاملة أكثر تنوعا ، ككبار السن ، وتتألف من المزيد من النساء ○ مخاوف رئيسة من الأمن الداخلي والهجمات الإرهابية ○ ضرورة قانون Sarbanes-Oxley وغيرها من التشريعات المرتبطة بالقانون ○ زيادة المسؤولية الاجتماعية للشركات ○ مزيد من التركيز على الاستدامة

Organizational Responses

- Be Reactive, Anticipative, Adaptive, and Proactive
- Managers may take actions, such as
  - Employ strategic planning.
  - Restructure business processes.
  - Improve corporate information systems. … more [in your book]
  - Use new and innovative business models.
  - Participate in business alliances.

الاستجابات التنظيمية

- يكون تفاعلي ، توقعي ، متكيف ، و إستباقي.
- المدراء بإمكانهم اتخاذ إجراءات مثل :
  - توظيف التخطيط الاستراتيجي.
  - إعادة هيكلة العمليات التجارية.
  - تحسين نظم المعلومات المؤسسية.
  - استخدام نماذج أعمال جديدة ومبتكرة.
  - المشاركة في تحالفات الأعمال.

Closing the Strategy Gap

One of the major objectives of computerized decision support is to facilitate closing the gap between the current performance of an organization and its desired performance, as expressed in its mission, objectives, and goals, and the strategy to achieve them.

إغلاق الفجوة الاستراتيجية

يتمثل أحد الأهداف الرئيسية لدعم اتخاذ القرار بالحاسب الآلي في تسهيل سد الفجوة بين الأداء الحالي للمؤسسة والأداء المرغوب فيه ، كما هو موضح في مهمتها وأهدافها وغايتها ، واستراتيجية تحقيقها.

Managerial Decision Making

- **Management** is a process by which organizational goals are achieved by using resources.
  - **Inputs:** resources
  - **Output:** attainment of goals
  - **Measure of success:** outputs / inputs
- Management ≈ Decision Making
- **Decision making**: selecting the best solution from two or more alternatives

اتخاذ القرارات الإدارية

- **الإدارة** هي عملية يتم من خلالها تحقيق الأهداف التنظيمية باستخدام الموارد.
  - **المدخلات:** الموارد
  - **الناتج:** تحقيق الأهداف
  - **قياس النجاح:** المخرجات / المدخلات
- الإدارة ≈ صنع القرار
- **صنع القرار**: اختيار الحل الأفضل من بديلين أو أكثر.

The Nature of Managers' Work (Mintzberg's 10 Managerial Roles)

Interpersonal 1. Figurehead 2. Leader 3. Liaison	**التعامل مع الأخرين** 1. الرئيس الصوري (يعني عنده منصب لكن ما عنده القدرة للسيطرة بالمنظمة) 1. القائد او المرشد 2. اتصال متبادل او ارتباط.
Informational 4. Monitor 5. Disseminator 6. Spokesperson	**المعلومات** 4 المراقب 5 الناشر 6 الناطق الرسمي
Decisional 7 Entrepreneur 8 Disturbance handler 9 Resource allocator 10 Negotiator	**المتعلق بالقرارات** 7 رجل الاعمال 8 معالج الاضطرابات او الإزعاجات 9 الموارد المخصصة 10 المفاوض

Decision-making Process

- Managers usually make decisions by following a four-step process (a.k.a. the scientific approach)
  1. Define the problem (or opportunity)
  2. Construct a model that describes the real-world problem.
  3. Identify possible solutions to the modeled problem and evaluate the solutions.
  4. Compare, choose, and recommend a potential solution to the problem.

عملية اتخاذ القرار

- عادةً ما يتخذ المديرون قراراتهم عن طريق اتباع عملية من أربع خطوات (مثل النهج العلمي)
  1. تحديد المشكلة (أو الفرصة)
  2. بناء نموذج يصف الشكل الحقيقي للمشكلة.
  3. تحديد الحلول الممكنة للمشكلة النموذجية وتقييم الحلول.
  4. قارن ، واختار ، واقترح حلًا محتملًا للمشكلة.

Information Systems Support for Decision Making

- Group communication and collaboration
- Improved data management
- Managing data warehouses and Big Data
- Analytical support
- Overcoming cognitive limits in processing and storing information
- Knowledge management
- Anywhere, anytime support

دعم نظم المعلومات لصنع القرار

- التواصل الجماعي والتعاون
- تحسين إدارة البيانات
- إدارة مستودعات البيانات والبيانات الكبيرة
- دعم تحليلي
- التغلب على الحدود المعرفية في معالجة وتخزين المعلومات
- إدارة المعرفة
- الدعم في أي مكان ، في أي وقت

An Early Decision Support Framework (by Gory and Scott-Morten, 1971)

	Type of Control		
Type of Decision	**Operational Control**	**Managerial Control**	**Strategic Planning**
Structured	1 Accounts receivable Accounts payable Order entry	2 Budget analysis Short-term forecasting Personnel reports Make-or-buy	3 Financial management Investment portfolio Warehouse location Distribution systems
Semistructured	4 Production scheduling Inventory control	5 Credit evaluation Budget preparation Plant layout Project scheduling Reward system design Inventory categorization	6 Building a new plant Mergers & acquisitions New product planning Compensation planning Quality assurance HR policies Inventory planning
Unstructured	7 Buying software Approving loans Operating a help desk Selecting a cover for a magazine	8 Negotiating Recruiting an executive Buying hardware Lobbying	9 R & D planning New tech. development Social responsibility planning

An Early Decision Support Framework

- Degree of Structuredness (Simon, 1977)
  - Decisions are classified as
    - Highly structured (a.k.a. programmed)
    - Semi-structured
    - Highly unstructured (i.e., nonprogrammed)
- Types of Control (Anthony, 1965)
  - Strategic planning (top-level, long-range)
  - Management control (tactical planning)
  - Operational control

إطار دعم القرار المبكر

- درجة الهيكلة (سيمون ، 1977)
  - تصنف القرارات باسم
    - منظم للغاية (مبرمج)
    - شبه منظم
    - غير منظم للغاية (أي غير مبرمج)
- أنواع السيطرة (أنتوني ، 1965)
  - التخطيط الاستراتيجي (المستوى الأعلى ، المدى البعيد)
  - السيطرة على الإدارة (التخطيط التكتيكي)
  - السيطرة على العمليات

The Concept of DSS

- **DSS** - interactive computer-based systems, which help decision makers utilize data and models to solve unstructured problems *(Gorry and Scott-Morton, 1971)*
- Decision support systems couple the intellectual resources of individuals with the capabilities of the computer to improve the quality of decisions.
- DS as an Umbrella Term
- Evolution of DS into Business Intelligence

مفهوم DSS

- **DSS** - انظمة تفاعلية تعتمد على الكمبيوتر ، تساعد صانعي القرار على استخدام البيانات والنماذج لحل المشكلات غير المنظمة (غوري وسكوت مورتون ، 1971)
- تقترن أنظمة دعم القرار بالموارد الفكرية للأفراد ذوي قدرات في الكمبيوتر لتحسين جودة القرارات.
- DS كمصطلح شامل
- تطور DS في ذكاء الأعمال

A Framework for Business Intelligence (BI)

- **BI** is an evolution of decision support concepts over time
  - Then: Executive Information System
  - Now: Everybody's Information System (BI)
- BI systems are enhanced with additional visualizations, alerts, and performance measurement capabilities
- The term BI emerged from industry

إطار لاستخبارات الأعمال (BI)

- **BI** هو تطور مفاهيم دعم القرار بمرور الوقت
  - ثم: نظام المعلومات التنفيذي
  - الآن: نظام معلومات الجميع (BI)
- يتم تحسين أنظمة BI مع تصورات إضافية ، وتنبيهات ، وقدرات قياس الأداء
- مصطلح BI خرج من الصناعة.

Definition of BI

- **BI** is an umbrella term that combines architectures, tools, databases, analytical tools, applications, and methodologies
- BI is a content-free expression, so it means different things to different people
- BI's major objective is to enable easy access to data (and models) to provide business managers with the ability to conduct analysis
- BI helps *transform* data, to information (and knowledge), to decisions, and finally to action

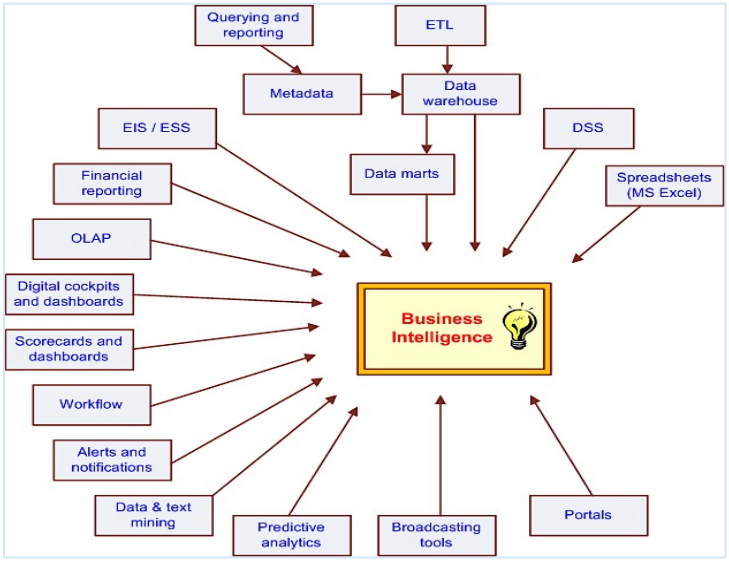
تعريف BI

- **BI** هو مصطلح شامل يجمع بين البنى والأدوات وقواعد البيانات والأدوات التحليلية والتطبيقات والمنهجيات
- **BI** هو تعبير خالٍ من المحتوى ، لذلك فهو يعني أشياء مختلفة لأشخاص مختلفين
- هدف BI الرئيسي هو تمكين الوصول السهل إلى البيانات (والنماذج) لتزويد مديري الأعمال بالقدرة على إجراء التحليل
- يساعد BI في تحويل البيانات والمعلومات (والمعرفة) ، إلى القرارات ، وفي النهاية إلى العمل

A Brief History of BI

- The term BI was coined by the Gartner Group in the mid-1990s
- However, the concept is much older
  - 1970s - MIS reporting - static/periodic reports
  - 1980s - Executive Information Systems (EIS)
  - 1990s - OLAP, dynamic, multidimensional, ad-hoc reporting -> coining of the term “BI”
  - 2010s - Inclusion of AI and Data/Text Mining capabilities; Web-based Portals/Dashboards, Big Data, Social Media, Analytics
  - 2020s - yet to be seen

The Evolution of BI Capabilities



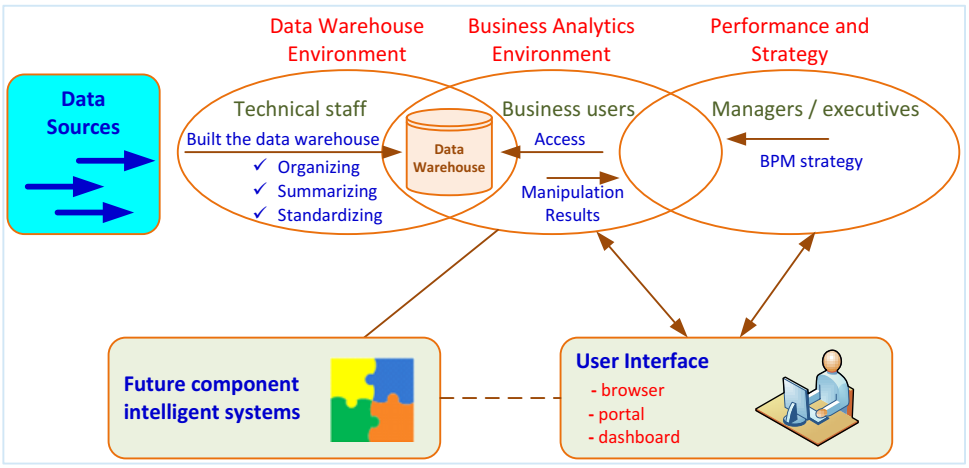
The Architecture of BI

- A BI system has four major components
  - a data warehouse, with its source data
  - business analytics, a collection of tools for manipulating, mining, and analyzing the data in the data warehouse
  - business performance management (BPM) for monitoring and analyzing performance
  - a user interface (e.g., dashboard)

هندسة BI

- يحتوي نظام BI على أربعة مكونات رئيسية
  - مستودع البيانات ، مع بيانات المصدر الخاصة به
  - تحليلات الأعمال التجارية ، وهي مجموعة من الأدوات للتلاعب والتعدين وتحليل البيانات في مستودع البيانات
  - إدارة أداء الأعمال (BPM) لرصد وتحليل الأداء
  - واجهة مستخدم (على سبيل المثال ، لوحة القيادة)

A High-Level Architecture of BI



Business Value of BI Analytical Applications	قيمة الاعمال لتحليل تطبيقات الBI
Customer segmentation	فئات الزبائن
Propensity to buy	الميل للشراء
Customer profitability	ربحية العميل
Fraud detection	الكشف عن الغش
Customer attrition	تناقص العملاء
Channel optimization	تحسين القناة

DSS-BI Connections

- Similarities and differences?
  - Similar architectures, data focus, …
- Direct vs. indirect support
- Different target audiences
- Commercially available systems versus in-house development of solutions
- Origination – Industry vs. Academia
- So, is DSS = BI ?

ترابط DSS-BI

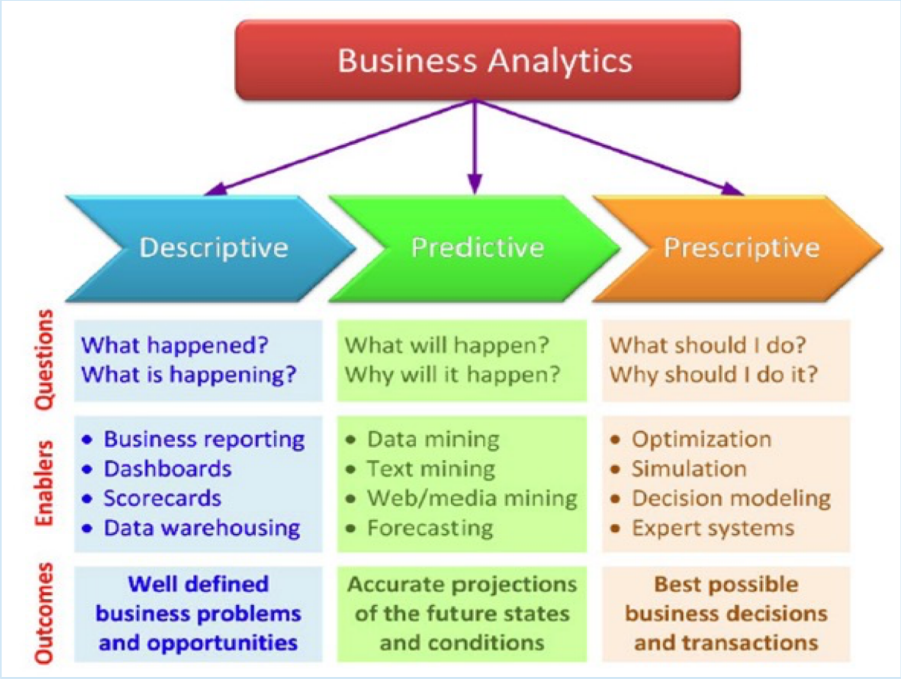
- أوجه التشابه و الاختلاف ؟
  - هندسة متماثله ، تركيز البيانات ..
- دعم مباشر vs دعم غير مباشر
- جمهور مستهدف مختلف
- الأنظمة المتاحة تجاريًا مقابل تطوير الحلول الداخلية
- نشأه – الصناعة vs الاكاديمي

Analytics Overview

- Analytics?
  - Something new or just a new name for …
- A Simple Taxonomy of Analytics (proposed by INFORMS)
  - Descriptive Analytics
  - Predictive Analytics
  - Prescriptive Analytics
- Analytics or Data Science?

نظرة عامه عن التحليلات

- تحليلات؟
  - شيء جديد أو مجرد اسم جديد ...
- تصنيف بسيط لبرنامج Analytics (مقترح من قبل INFORMS)
  - تحليلات وصفية
  - التحليلات التنبؤية
  - التحليلات المسبقة
- التحليلات أو علوم البيانات؟



Introduction to Big Data Analytics

- Big Data?
  - Not just big!
  - Volume
  - Variety
  - Velocity

More of Big Data and related analytics tools and techniques are covered in Chapter 13.

مقدمة لتحليلات البيانات الكبيرة

- البيانات الكبيرة ؟
  - ليست فقط كبيره!
  - حجم
  - نوع
  - السرعة

Foundations and Technologies for Decision Making

Characteristics of Decision Making

- Groupthink
- Evaluating what-if scenarios
- Experimentation with a real system!
- Changes in the decision-making environment may occur continuously
- Time pressure on the decision maker
- Analyzing a problem takes time/money
- Insufficient or too much information

خصائص صنع القرار

- التفكير الجماعي
- تقييم سيناريوهات "ماذا لو"
- التجريب مع نظام حقيقي!
- قد تحدث التغييرات في بيئة اتخاذ القرار بشكل مستمر
- ضغط الوقت على صانع القرار
- تحليل المشكلة يتطلب الوقت / المال
- ناقص أو الكثير من المعلومات

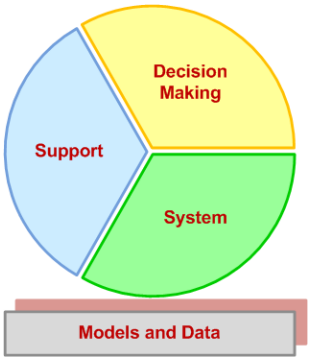
Characteristics of Decision Making - Decision Support Systems (DSS)

Dissecting DSS into its main concepts →

Building successful DSS requires a thorough understanding of these concepts

شرح DSS في مفاهيمها الرئيسية

لبناء DSS بشكل ناجحًا! يتطلب فهمًا شاملاً لهذه المفاهيم.



Decision Making

- A process of choosing among two or more alternative courses of action for the purpose of attaining a goal(s)
- **Managerial decision making** is synonymous with the entire management process - *Simon (1977)*
- Example: Planning
  - What should be done? When? Where? Why? How? By whom?

صنع القرار

- عملية اختيار من بين مسارين بديلين أو أكثر للعمل من أجل تحقيق هدف (أهداف)
- **صنع القرار الإداري** مرادف لعملية الإدارة بأكملها - Simon (1977)
- مثال: التخطيط
  - ما الذي يجب إتمامه؟ متي؟ أين؟ لماذا ا؟ ماذا؟ بواسطة من؟

Decision-Making Disciplines

- **Behavioral:** anthropology, law, philosophy, political science, psychology, social psychology, and sociology
- **Scientific:** computer science, decision analysis, economics, engineering, the hard sciences (e.g., biology, chemistry, physics), management science/operations research, mathematics, and statistics
- Each discipline has its own set of assumptions and each contributes a unique, valid view of how people make decisions

تخصصات صنع القرار

- **سلوكياً:** الأنثروبولوجيا والقانون والفلسفة والعلوم السياسية وعلم النفس وعلم النفس الاجتماعي وعلم الاجتماع
- **علميًا:** علوم الكمبيوتر وتحليل القرارات والاقتصاد والهندسة والعلوم الصعبة (مثل علم الأحياء والكيمياء والفيزياء) وأبحاث إدارة العلوم والعمليات والرياضيات والإحصاء
- لكل تخصص مجموعة خاصة به من الافتراضات ويساهم كل منها في عرض فريد وصالح لكيفية اتخاذ الناس للقرارات

Decision-Making Disciplines (cont.)	تابع – تخصصات صنع القرار
❖ Better decisions ○ Tradeoff: **accuracy** versus **speed** ❖ Fast decision may be detrimental ❖ Many areas suffer from fast decisions ❖ Effectiveness versus Efficiency ❖ Effectiveness → “goodness” “accuracy” ❖ Efficiency → “speed” “less resources” ❖ A fine balance is what is needed!	❖ قرارات أفضل ○ المقايضة: **الدقة** مقابل **السرعة** ❖ قد يكون القرار السريع ضارًا ❖ العديد من المجالات تعاني من قرارات سريعة ❖ الفعالية مقابل الكفاءة ❖ فعالية → "الجودة" "الدقة" ❖ الكفاءة → "السرعة" "أقل الموارد" ❖ التوازن الدقيق هو ما نحتاجه!

Decision Style

- The manner by which decision makers think and react to problems
  - perceive a problem
  - cognitive response
  - values and beliefs
- When making decisions, people…
  - follow different steps/sequence
  - give different emphasis, time allotment, and priority to each step

أسلوب القرارات

- الأسلوب الذي يفكروا به صُناع القرار و الاستجابة مع المشاكل
  - يدرك مشكلة
  - الاستجابة المعرفية
  - القيم والمعتقدات
- عند اتخاذ القرارات ، الناس ...
  - اتبع الخطوات / التسلسل المختلفة
  - إعطاء تركيز مختلف ، وتخصيص الوقت ، والأولوية لكل خطوة

- Personality temperament tests are often used to determine decision styles
- **There are many such tests**
  - Meyers/Briggs,
  - True Colors (Birkman),
  - Keirsey Temperament Theory, …
- Various tests measure somewhat different aspects of personality
  - They cannot be equated!

- غالبًا ما تستخدم اختبارات طبع الشخصية لتحديد أنماط القرار
- هناك العديد من هذه الاختبارات
  - Meyers/Briggs ،
  - ألوان حقيقية (بيركمان) ،
  - نظرية Keirsey Temperament ،...
- اختبارات مختلفة لقياس جوانب مختلفة إلى حد ما من الشخصية
  - لا يمكن ان تكون متساوية

- **Decision-making styles**
  - Heuristic versus Analytic
  - Autocratic versus Democratic
  - Consultative (with individuals or groups)
- A successful computerized system should fit the decision style and the decision situation
  - Should be flexible and adaptable to different users (individuals vs. groups)

- **أساليب صنع القرار**
  - ارشادي vs التحليلي
  - مستبد او المتسلط vs الديمقراطية
  - استشاري (مع أفراد أو مجموعات)
- يجب أن يتلاءم نظام الكمبيوتر الناجح مع أسلوب القرار وحالة القرار
  - يجب أن تكون مرنة وقابلة للتكيف مع مختلف المستخدمين (الأفراد vs المجموعات)

Decision Makers

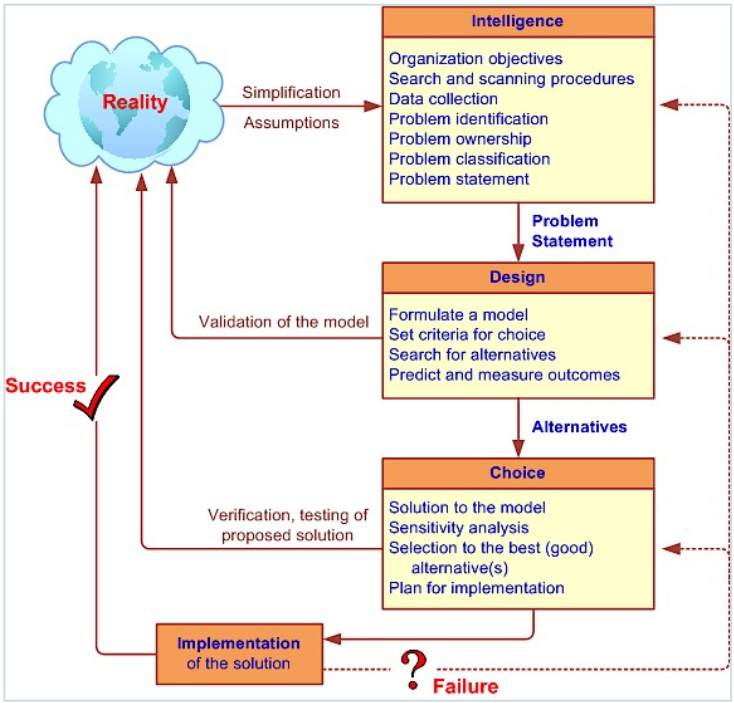
- **Small** organizations
  - Individuals
  - Conflicting objectives
- **Medium-to-large** organizations
  - Groups
  - Different styles, backgrounds, expectations
  - Conflicting objectives
  - Consensus is often difficult to reach
  - Help: Computer support, GSS, …

صُنّاع القرار

- المنظمات **الصغيرة**
  - الأفراد
  - أهداف متعارضه
- المؤسسات **المتوسطة إلى الكبيرة**
  - المجموعات
  - أنماط مختلفة ، خلفيات ، توقعات
  - أهداف متعارضه
  - من الصعب في الغالب الوصول إلى توافق في الآراء
  - مساعدة: دعم الكمبيوتر ، GSS ، ...

Phases of Decision-Making Process	مراحل عملية صنع القرار
Humans consciously or subconsciously follow a **systematic decision-making process** - Simon (1977) 1) Intelligence 2) Design 3) Choice 4) Implementation 5) (?) Monitoring (a part of intelligence?)	البشر بوعي أو لا شعوريًا يتبعون **عملية صنع قرار بشكل منظم** - سيمون (1977) .1 الذكاء .2 التصميم .3 خيار .4 التنفيذ .5 مراقبة (جزء من الذكاء؟)

Simon's Decision-Making Process



Decision Making: Intelligence Phase

- Scan the environment, either intermittently or continuously
- Identify problem situations or opportunities
- Monitor the results of the implementation
- **Problem** is the difference between what people desire (or expect) and what is actually occurring
  - Symptom versus Problem
- Timely identification of opportunities is as important as identification of problems

صنع القرار: مرحلة الذكاء

- تفحص البيئة ، سواء بشكل متقطع أو مستمر
- تحديد حالات المشاكل أو الفرص
- مراقبة نتائج التنفيذ
- **المشكلة** هي الفرق بين ما يرغب فيه الناس (أو يتوقعونه) وما يحدث بالفعل
  - أعراض vs مشكلة
- تحديد الفرص في الوقت المناسب لا يقل أهمية عن تحديد المشاكل

- Potential issues in data/information collection and estimation
  - Lack of data
  - Cost of data collection
  - Inaccurate and/or imprecise data
  - Data estimation is often subjective
  - Data may be insecure
  - Key data may be qualitative
  - Data change over time (time-dependence)

- القضايا المحتملة في جمع البيانات والمعلومات وتقديرها
  - نقص البيانات
  - تكلفة جمع البيانات
  - بيانات خاطئة و/أو غير دقيقة
  - غالبًا ما يكون تقدير البيانات ذاتيًا
  - قد تكون البيانات غير آمنة
  - البيانات الرئيسية قد تكون نوعية
  - تغيير البيانات بمرور الوقت (الاعتماد على الوقت)

- Problem Classification
  - Classification of problems according to the degree of structuredness
- Problem Decomposition
  - Often solving the simpler subproblems may help in solving a complex problem.
  - Information/data can improve the structuredness of a problem situation
- Problem Ownership
- Outcome of intelligence phase → **A Formal Problem Statement**

- تصنيف المشكلة
  - تصنيف المشاكل حسب درجة الهيكلة
- تحلل المشكلة
  - غالباً ما يساعد حل المشاكل الفرعية البسيطة في حل مشكلة معقدة.
  - المعلومات / البيانات يمكن أن تحسن هيكلة حالة المشكلة
- تملك المُشكلة
- نتيجة مرحلة الذكاء → **بيان مشكلة رسمي**

Web and the Decision-Making Process

TABLE 2.1 Simon's Four Phases of Decision Making and the Web

Phase	Web Impacts	Impacts on the Web
Intelligence	Access to information to identify problems and opportunities from internal and external data sources	Identification of opportunities for e-commerce, Web infrastructure, hardware and software tools, etc.
	Access to analytics methods to identify opportunities	Intelligent agents, which reduce the burden of information overload
	Collaboration through group support systems (GSS) and knowledge management systems (KMS)	Smart search engines
Design	Access to data, models, and solution methods	Brainstorming methods (e.g., GSS) to collaborate in Web infrastructure design
	Use of online analytical processing (OLAP), data mining, and data warehouses	Models and solutions of Web infrastructure issues
	Collaboration through GSS and KMS	
	Similar solutions available from KMS	
Choice	Access to methods to evaluate the impacts of proposed solutions	Decision support system (DSS) tools, which examine and establish criteria from models to determine Web, intranet, and extranet infrastructure
		DSS tools, which determine how to route messages
Implementation	Web-based collaboration tools (e.g., GSS) and KMS, which can assist in implementing decisions	Decisions implemented on browser and server design and access, which ultimately determined how to set up the various components that have evolved into the Internet
	Tools, which monitor the performance of e-commerce and other sites, including intranets, extranets, and the Internet	

المرحلة	تأثير الويب	التأثيرات على الويب
الذكاء	❖ الوصول إلى المعلومات لتحديد المشاكل والفرص من مصادر البيانات الداخلية والخارجية. ❖ الوصول إلى أساليب التحليلات لتحديد الفرص. ❖ التعاون من خلال نظام دعم المجموعة (GSS) ونظام إدارة المعرفة (KMS).	❖ تحديد فرص التجارة الإلكترونية والبنية التحتية على شبكة الإنترنت وأدوات الأجهزة والبرامج ، إلخ. ❖ العوامل الذكية ، والتي تقلل من عبء المعلومات الزائد. ❖ محركات البحث الذكية.
التصميم	❖ الوصول إلى البيانات والنماذج وطرق الحل. ❖ استخدام المعالجة التحليلية عبر الإنترنت (OLAP)، استخراج البيانات ، ومستودعات البيانات ❖ التعاون من خلال GSS و KMS. ❖ حلول مماثلة متاحة من KMS.	❖ أساليب العصف الذهني (على سبيل المثال ، GSS) للتعاون في تصميم البنية الأساسية للويب ❖ نماذج وحلول قضايا البنية التحتية على شبكة الإنترنت
الاختيار	❖ الوصول إلى طرق لتقييم تأثيرات الحلول المقترحة.	❖ أدوات نظام دعم القرار(DSS)، التي تؤسس وتفحص المعايير لنماذج لتحديد البنية التحتية على شبكة الإنترنت والإنترانت والإكسترانت. ❖ أدوات DSS ، التي تحدد كيفية توجيه الرسائل
التنفيذ	❖ أدوات التعاون القائمة على الويب مثل (GSS) و KMS ، والتي يمكن أن تساعد في تنفيذ القرارات ❖ الأداة التي تراقب أداء مواقع التجارة الإلكترونية وغيرها ، بما في ذلك الشبكات الداخلية والخارجية والإنترنت	❖ تنفيذ القرار على تصميم المستعرض والخادم والوصول إليه ، والذي يحدد في النهاية كيفية إعداد المكونات المختلفة التي تطورت إلى الإنترنت.

Decision Making: The Design Phase

- Finding/developing and analyzing possible courses of actions
- A model of the decision-making problem is constructed, tested, and validated
- **Modeling**: conceptualizing a problem and abstracting it into a quantitative and/or qualitative form (i.e., using symbols/variables)
  - **Abstraction:** making assumptions for simplification
  - **Tradeoff (cost/benefit):** more or less abstraction
  - **Modeling:** both an art and a science

صنع القرار: مرحلة التصميم

- إيجاد / تطوير وتحليل مسارات الإجراءات الممكنة
- يتم إنشاء نموذج لمشكلة اتخاذ القرار واختباره والتحقق من صحته
- **النمذجة**: تصور مشكلة ما وتجريدها إلى شكل كمي و/ أو نوعي (أي باستخدام الرموز / المتغيرات)
  - **التجريد**: وضع افتراضات للتبسيط
  - **المقايضة (التكلفة / الفائدة):** تجريد أكثر أو أقل
  - **النمذجة:** كل من الفن والعلوم

- Selection of a **Principle of Choice**
  - It is a criterion that describes the acceptability of a solution approach
  - Reflection of decision-making objective(s)
  - In a model, it is the result variable
  - Choosing and validating against
    - High-risk versus low-risk
    - Optimize versus satisfice
  - Criterion is not a constraint!

- اختيار **مبدأ الاختيار**
  - وهو معيار يصف قبول نهج الحل
  - انعكاس هدف (أهداف) صنع القرار
  - في النموذج ، هو متغير النتيجة
  - اختيار والتحقق من صحة الضد
    - عالية المخاطر vs مخاطر منخفضة
    - تحسين vs الرضى
  - المعيار ليس عائقا!

- Normative models (= optimization)
  - the chosen alternative is demonstrably the best of all possible alternatives
  - Assumptions of rational decision makers
    - Humans are economic beings whose objective is to maximize the attainment of goals
    - For a decision-making situation, all alternative courses of action and consequences are known
    - Decision makers have an order or preference that enables them to rank the desirability of all consequences

- النماذج المعيارية (= التحسين)
  - البديل المختار هو بشكل واضح أفضل كل البدائل الممكنة
  - افتراضات صانعي القرار العقلاني
    - البشر هم الكائنات الاقتصادية التي تهدف إلى تحقيق أقصى قدر من تحقيق الأهداف
    - بالنسبة لحالة اتخاذ القرار ، تُعرف جميع مسارات العمل والعواقب البديلة
    - لدى صانعي القرار أمر أو تفضيل يمكنهم من ترتيب استحسان جميع العواقب

- Heuristic models (= sub-optimization)
  - The chosen alternative is the best of only a subset of possible alternatives
  - Often, it is not feasible to optimize realistic (size/complexity) problems
  - Sub-optimization may also help relax unrealistic assumptions in models
  - Help reach a good enough solution faster

- نماذج إرشادية (= التحسين الفرعي)
  - البديل المختار هو أفضل مجموعة فرعية فقط من البدائل الممكنة
  - في كثير من الأحيان ، ليس من الممكن تحسين المشاكل الواقعية (الحجم / التعقيد)
  - قد يساعد أيضًا الاستبعاد الفرعي في تخفيف الافتراضات غير الواقعية في النماذج
  - مساعدة في التوصل إلى حل جيد بما فيه الكفاية بشكل أسرع

Decision Making: The Design Phase (cont.)

- Descriptive models
  - Describe things as they are or as they are believed to be (mathematically based)
  - They do not provide a solution but information that may lead to a solution
  - **Simulation** - most common descriptive modeling method (mathematical depiction of systems in a computer environment)
  - Allows experimentation with the descriptive model of a system

- نماذج وصفية
  - وصف الأشياء كما هي أو كما يعتقد أنها (تستند إلى الرياضيات)
  - فهي لا تقدم حلاً ولكن المعلومات التي قد تؤدي إلى حل
  - **المحاكاة** - طريقة النمذجة الوصفية الأكثر شيوعًا (التصوير الرياضي للأنظمة في بيئة الكمبيوتر)
  - يسمح بالتجريب مع النموذج الوصفي للنظام
- Good Enough, or Satisficing
  - "something less than the best"
  - A form of sub-optimization
  - Seeking to achieve a desired level of performance as opposed to the "best"
  - Benefit: time saving
  - Simon's idea of bounded rationality

- جيد بما فيه الكفاية ، أو الرضا
  - شكل من أشكال التحسين الفرعي
  - السعي إلى تحقيق مستوى الأداء المطلوب بدلاً من "الأفضل"
  - فائدة: توفير الوقت
  - فكرة سيمون عن العقلانية المحدودة
- Developing (Generating) Alternatives
  - In optimization models (such as linear programming), the alternatives may be generated automatically
  - In most MSS situations, however, it is necessary to generate alternatives manually
  - Use of GSS helps generate alternatives
- Measuring/ranking the outcomes
  - Using the principle of choice

- تطوير (انتاج) البدائل
  - في نماذج التحسين (مثل البرمجة الخطية) ، قد يتم إنشاء البدائل تلقائيًا
  - ومع ذلك ، في معظم حالات MSS، من الضروري توليد البدائل يدوياً
  - يساعد استخدام GSS في إنشاء بدائل
- قياس / ترتيب النتائج
  - باستخدام مبدأ الاختيار
- Risk
  - Lack of precise knowledge (uncertainty)
  - Risk can be measured with probability
- Scenario (what-if case)
  - A statement of assumptions about the operating environment (variables) of a particular system at a given time
  - **Possible scenarios:** best, worst, most likely, average (and custom intervals)

- الخطر
  - نقص المعرفة الدقيقة (عدم اليقين)
  - يمكن قياس المخاطر باحتمالية
- السيناريو (ماذا لو كانت القضية)
  - بيان الافتراضات حول بيئة التشغيل (المتغيرات) لنظام معين في وقت معين
  - **السيناريوهات المحتملة:** الأفضل والأسوأ والأرجح والمتوسط (والفواصل الزمنية المخصصة)

Decision Making: The Choice Phase

- The actual decision and the commitment to follow a certain course of action are made here
- The boundary between the design and choice is often unclear (partially overlapping phases)
  - Generate alternatives while performing evaluations
- Includes the search, evaluation, and recommendation of an appropriate solution to the model
- Solving the model versus solving the problem!

صنع القرار: مرحلة الاختيار

- يتم اتخاذ القرار الفعلي والالتزام باتباع نهج معين من العمل هنا
- الحدود بين التصميم والاختيار غالبًا ما تكون غير واضحة (مراحل متداخلة جزئيًا)
  - توليد بدائل أثناء إجراء التقييمات
- يتضمن البحث والتقييم والتوصية بالحل المناسب للنموذج
- حل النموذج vs حل المشكلة!

Search approaches ○ Analytic techniques (solving with a formula) ○ Algorithms (step-by-step procedures) ○ Heuristics (rule of thumb) ○ Blind search (truly random search)	**مناهج البحث** ○ التقنيات التحليلية (الحل مع الصيغة) ○ الخوارزميات (إجراءات خطوة بخطوة) ○ الاستدلال (حكم مبني على تجربة) ○ بحث أعمى او مموهـ (بحث عشوائي حقا)
Additional activities ○ Sensitivity analysis ○ What-if analysis ○ Goal seeking	**أنشطة إضافية** ○ تحليل الحساسية ○ تحليل ماذا لو ○ البحث عن الهدف

Decision Making: The Implementation Phase

"Nothing more difficult to carry out, nor more doubtful of success, nor more dangerous to handle, than to initiate a new order of things." - *The Prince, Machiavelli 1500s*

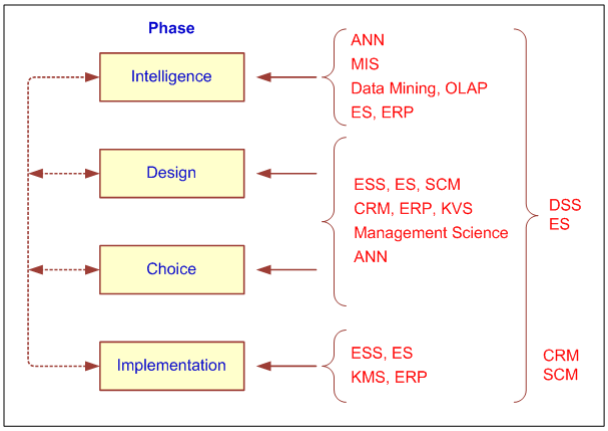
- Solution to a problem → Change
- Change management ?..
- **Implementation**: putting a recommended solution to work

صنع القرار: مرحلة التنفيذ

"لا شيء يصعب تنفيذه ، ولا شك في النجاح ، ولا أكثر خطورة في التعامل معه ، بدلاً من بدء نظام جديد للأشياء". - الأمير ، ميكافيلي s1500

- حل المشكلة → تغيير
- تغيير الإدارة ؟..
- **التنفيذ**: وضع حل موصى به للعمل.

How Decisions are Supported



How Decisions are Supported (cont.)

- Support for the Intelligence Phase
  - Enabling continuous scanning of external and internal information sources to identify problems and/or opportunities
  - Resources/technologies: Web; ES, OLAP, data warehousing, data/text/Web mining, EIS/Dashboards, KMS, GSS, GIS,…
  - **B**usiness **a**ctivity **m**onitoring (**BAM**)
  - **B**usiness **p**rocess **m**anagement (**BPM**)
  - **P**roduct **l**ife-cycle **m**anagement (**PLM**)

كيف يتم دعم القرارات

- دعم لمرحلة الذكاء
  - تمكين المسح المستمر لمصادر المعلومات الخارجية والداخلية لتحديد المشكلات و/أو الفرص
  - الموارد/ التقنيات: Web; ES, OLAP, data warehousing, data/text/Web mining, EIS/Dashboards, KMS, GSS, GIS,…
  - مراقبة النشاط التجاري (**BAM**)
  - إدارة العمليات التجارية (**BPM**)
  - إدارة دورة حياة المنتج (**PLM**)

- Support for the Design Phase
  - Enabling generating alternative courses of action, determining the criteria for choice
  - Generating alternatives
    - Structured/simple problems: standard and/or special models
    - Unstructured/complex problems: human experts, ES, KMS, brainstorming/GSS, OLAP, data/text mining
- A good “criteria for choice” is critical!

- دعم مرحلة التصميم
  - تمكين توليد مسارات بديلة للعمل ، وتحديد معايير الاختيار
  - توليد البدائل
    - هيكلية او تشييد/المشاكل البسيطة: النماذج القياسية و/ أو الخاصة
    - المشكلات غير المعقدة / المعقدة: الخبراء البشريون ، ES ، KMS ، brainstorming/GSS ، OLAP ، data/text mining
- تعتبر "معايير الاختيار" الجيدة أمرًا بالغ الأهمية!

- Support for the Choice Phase
  - Enabling selection of the best alternative given a complex constraint structure
  - Use sensitivity analyses, what-if analyses, goal seeking
  - Resources
    - KMS
    - CRM, ERP, and SCM
    - Simulation and other descriptive models

- دعم مرحلة الاختيار
  - تمكين اختيار البديل الأفضل في ضوء هيكل القيد المعقد
  - استخدام تحليلات الحساسية ، تحليلات ماذا لو ، البحث عن الأهداف
  - المصادر
    - KMS
    - CRM و ERP و SCM
    - المحاكاة والنماذج الوصفية الأخرى

- Support for the Implementation Phase
  - Enabling implementation/deployment of the selected solution to the system
  - Decision communication, explanation and justification to reduce resistance to change
  - Resources
    - Corporate portals, Web 2.0/Wikis
    - Brainstorming/GSS
    - KMS, ES

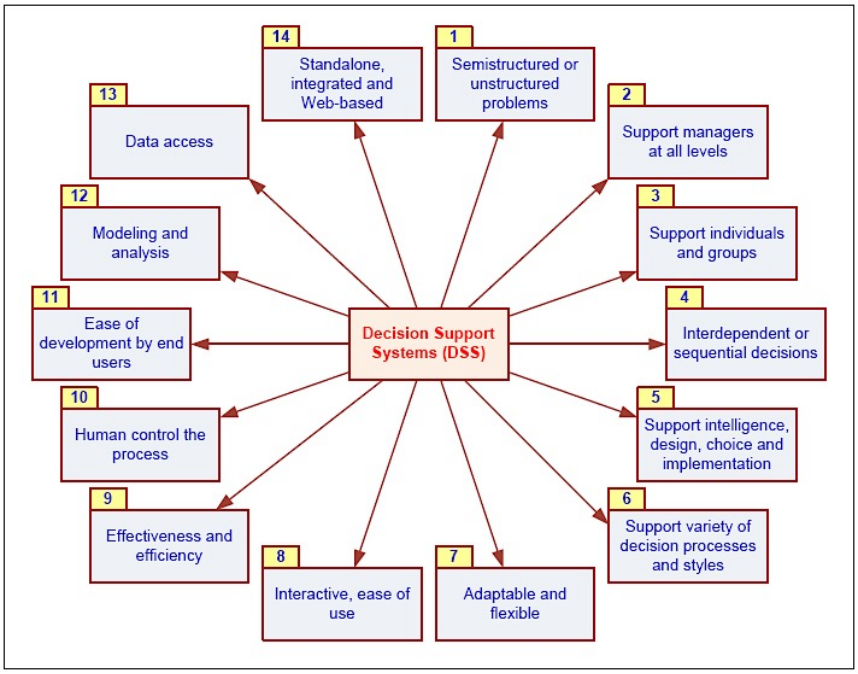
- دعم مرحلة التنفيذ
  - تمكين تنفيذ / نشر الحل المحدد للنظام
  - الاتصال او تبادل القرار ، والشرح والتبرير للحد من مقاومة التغيير
  - مصادر
    - مداخل الشركه ، Web 2.0 / Wikis
    - العصف الذهني / GSS
    - ES ،KMS

DSS Capabilities

- **DSS early definition:** it is a system intended to support managerial decisions in semi-structured and unstructured decision situations
- DSS were meant to be adjuncts to decision makers → extending their capabilities
- They are computer based and would operate interactively online, and preferably would have graphical output capabilities
- Nowadays, simplified via Web browsers and mobile devices

قدرات DSS

- **DSS تعريف مبكر:** إنه نظام يهدف إلى دعم القرارات الإدارية في حالات القرار شبه المنظم وغير المنظم
- كان المقصود من نظام دعم القرارات أن يكون ملحقًا لصانعي القرار → توسيع قدراتهم
- فهي تستند إلى الكمبيوتر وتعمل بشكل تفاعلي عبر الإنترنت ، ويفضل أن يكون لديها إمكانات إخراج رسومية
- في الوقت الحاضر ، مبسطة عبر متصفحات الويب والأجهزة المحمولة



DSS Classifications

- AIS SIGDSS Classification
  1. Communication-driven and group DSS
  2. Data-driven DSS
  3. Document-driven DSS
  4. Knowledge-driven DSS
  5. Model-driven DSS
- Often DSS is a hybrid of many classes

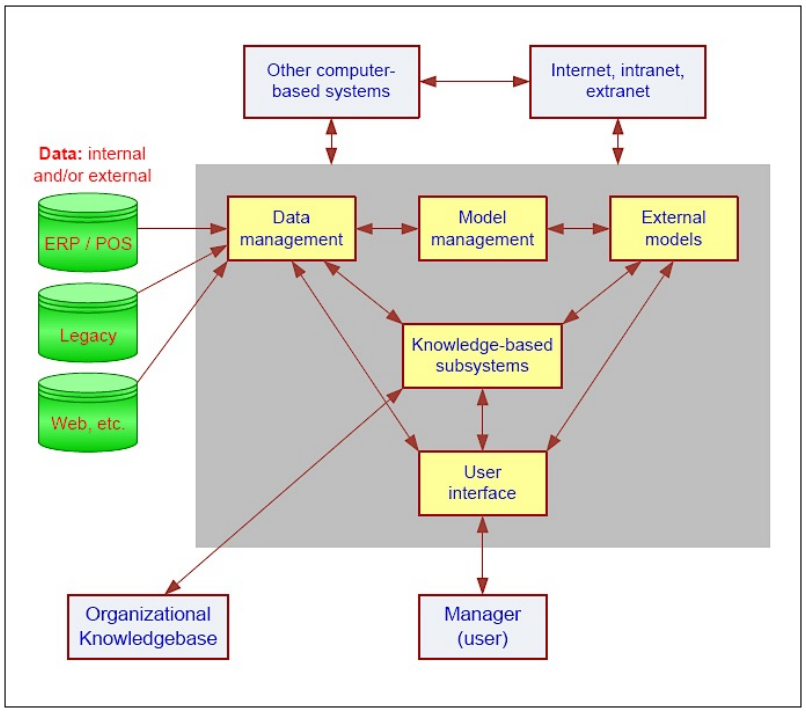
تصنيفات DSS

- تصنيف AIS SIGDSS
  1. قيام الاتصالات و قروب على DSS
  2. البيانات مستندة الى DSS
  3. الوثائق مستندة الى DSS
  4. المعرفة مستندة الى DSS
  5. النموذج مستند الى DSS
- في كثير من الأحيان DSS هو هجين من العديد من الطبقات

- Other **DSS Categories**
  - Institutional and ad-hoc DSS
  - Custom-made systems versus ready-made systems
  - Personal, group, and organizational support
  - Individual support system versus group support system (GSS)…

- **فئات DSS** الأخرى
  - DSS المؤسسية والمخصصة
  - أنظمة مصنوعة خصيصا vs أنظمة جاهزة
  - الدعم الشخصي والجماعي والتنظيمي
  - نظام الدعم الفردي vs نظام دعم المجموعة (GSS) ...

Components of DSS



Components of DSS (cont.)

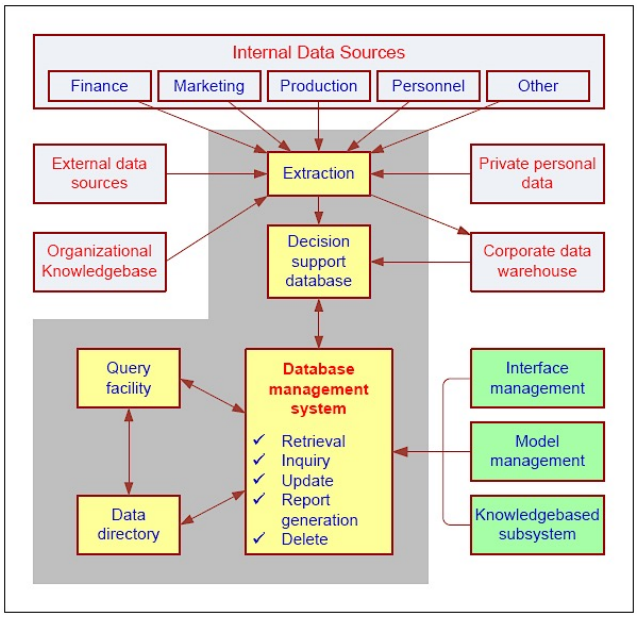
1. Data Management Subsystem
   - Includes the database that contains the data
   - Database management system (DBMS)
   - Can be connected to a data warehouse
2. Model Management Subsystem
   - Model base management system (MBMS)
3. User Interface Subsystem
4. Knowledgebase Management Subsystem
   - Organizational knowledge base

مكونات DSS

1. نظام إدارة البيانات الفرعي
   - يتضمن قاعدة البيانات التي تحتوي على البيانات
   - نظام إدارة قواعد البيانات (DBMS)
   - يمكن أن تكون متصلا بمخزن البيانات
2. النظام الفرعي لإدارة النماذج
   - نظام إدارة القواعد النموذجي (MBMS)
3. نظام الفرعي لواجهة المستخدم
4. النظام الفرعي لإدارة المعرفة
   - قاعدة المعرفة التنظيمية

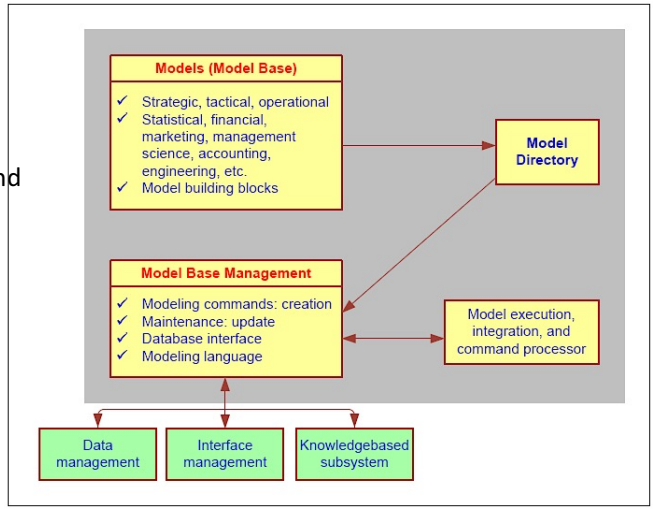
DSS Components: Data Management Subsystem

- DSS database
- DBMS
- Data directory
- Query facility



DSS Components: Model Management Subsystem

- Model base
- MBMS
- Modeling language
- Model directory
- Model execution, integration, and command processor

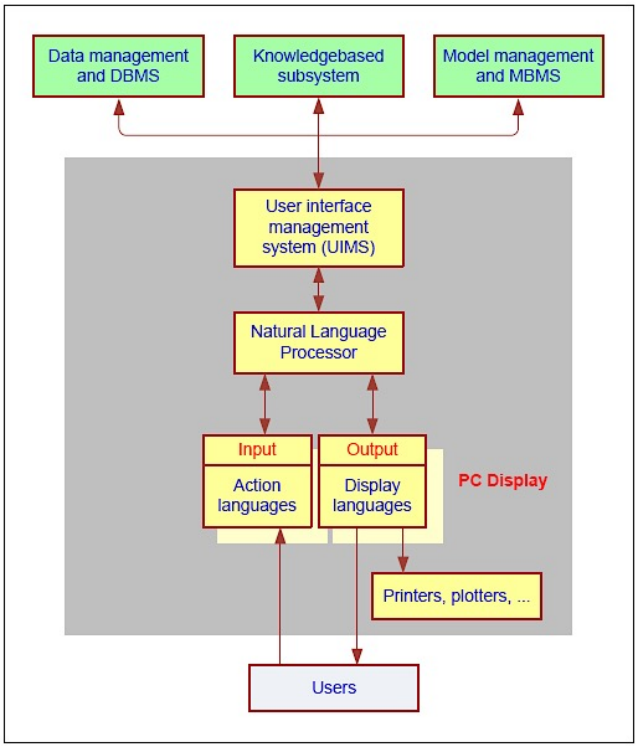


DSS Components: User Interface Subsystem

- Interface
  - Application interface
  - User Interface (GUI?)
- DSS User Interface
  - Portal
  - Graphical icons
    - Dashboard
  - Color coding
- Interfacing with PDAs, cell phones, etc.
  - See Technology Insight 2.2 for next gen devices

عناصر DSS – الانظمة الفرعية لواجهة المستخدم

- الواجهة
  - واجهة التطبيق
  - واجهة المستخدم (واجهة المستخدم الرسومية؟)
- واجهة مستخدم DSS
  - بوابة
  - الرموز الرسومية
    - لوحة القيادة
  - لون الترميز
- التواصل مع (PDA) والهواتف الخلوية وما إلى ذلك



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

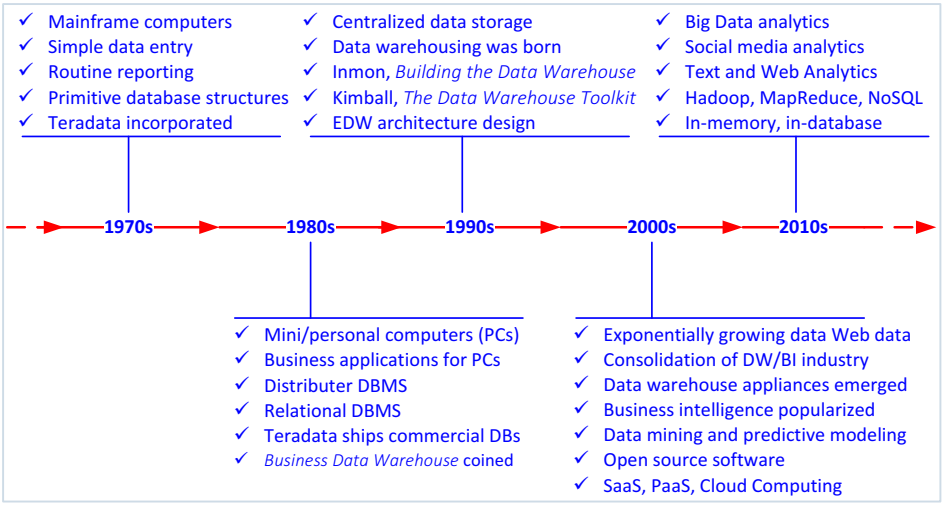
What is a Data Warehouse?

- 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

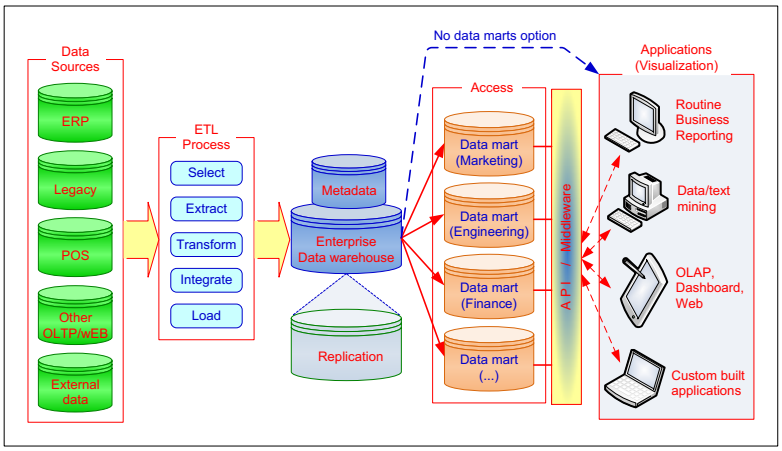
- **Operational data stores (ODS)**
  - A type of database often used as an interim area for a data warehouse
- **Oper marts** - an operational data mart.
- **Enterprise data warehouse (EDW)**
  - A data warehouse for the enterprise.
- **Metadata**: Data about data.
  - In a data warehouse, metadata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



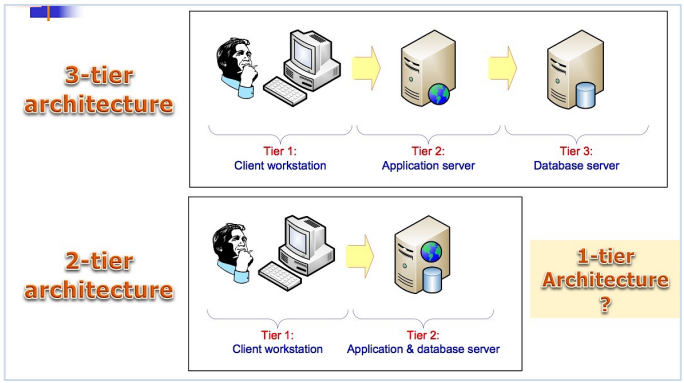
DW Architecture

- 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؟



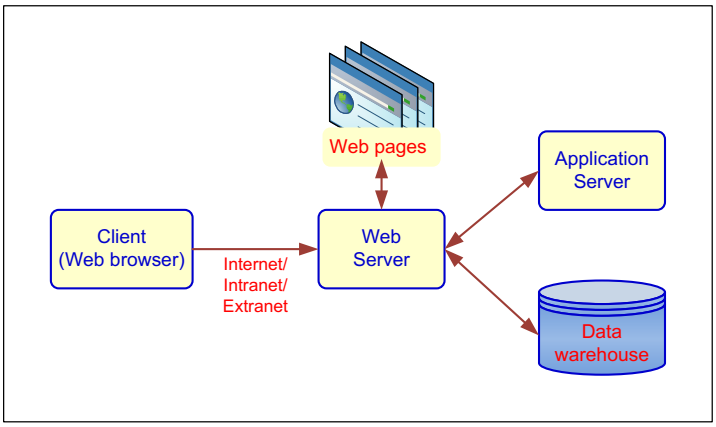
Data Warehousing Architectures

- 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?

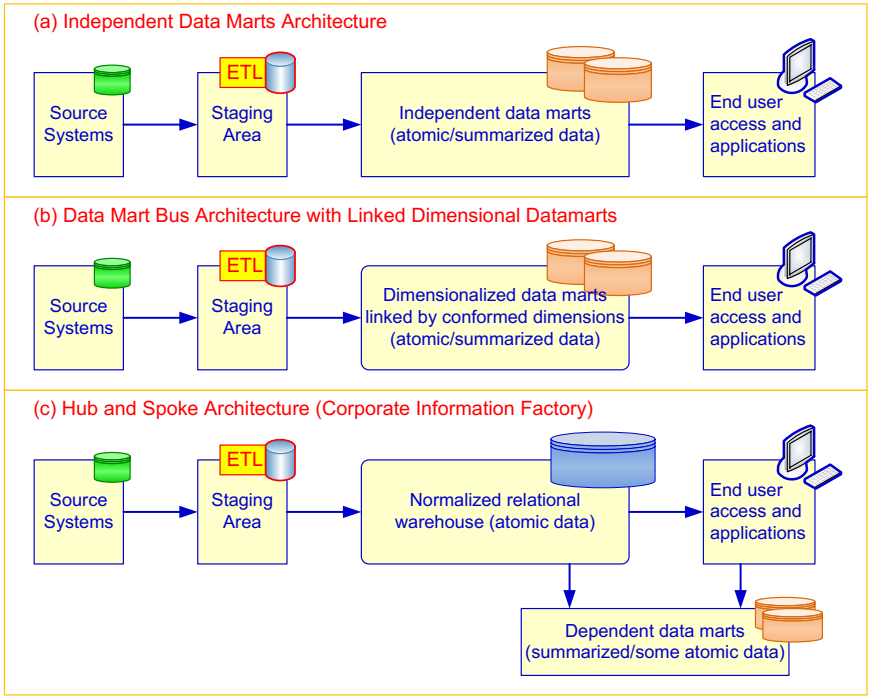
هياكل تخزين البيانات

- 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؟

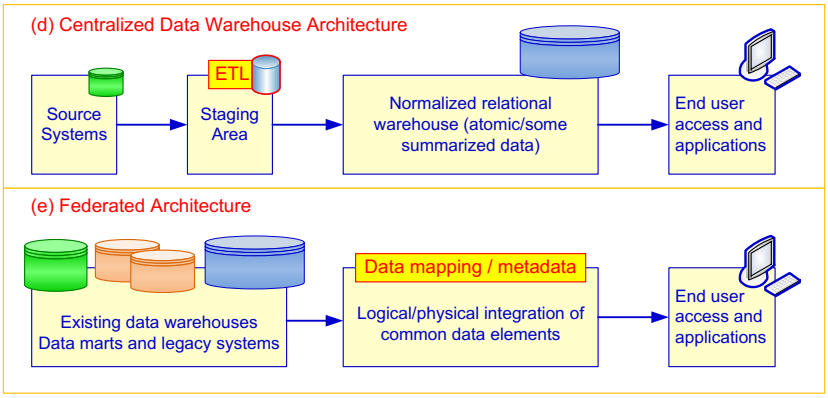
A Web-Based DW Architecture



Alternative DW Architectures



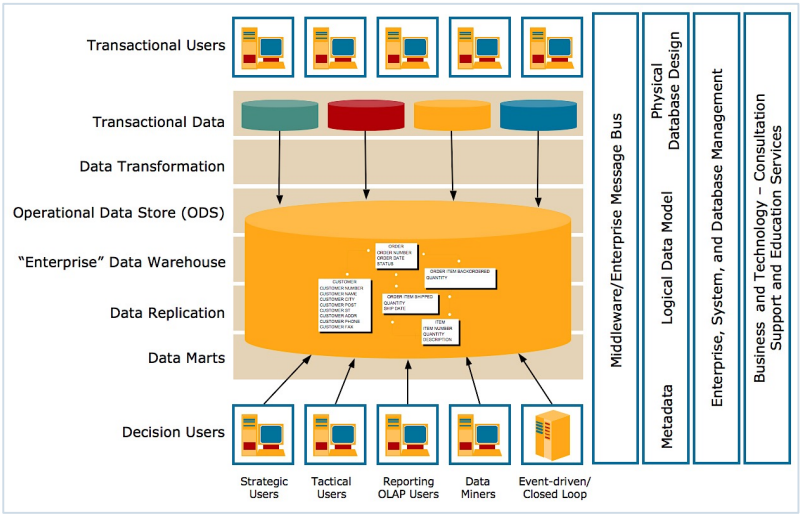
Alternative DW Architectures (cont.)



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

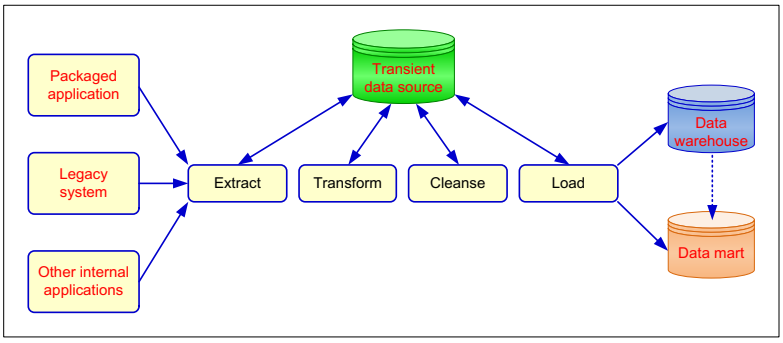


Data Integration and the Extraction, Transformation, and Load Process

- **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

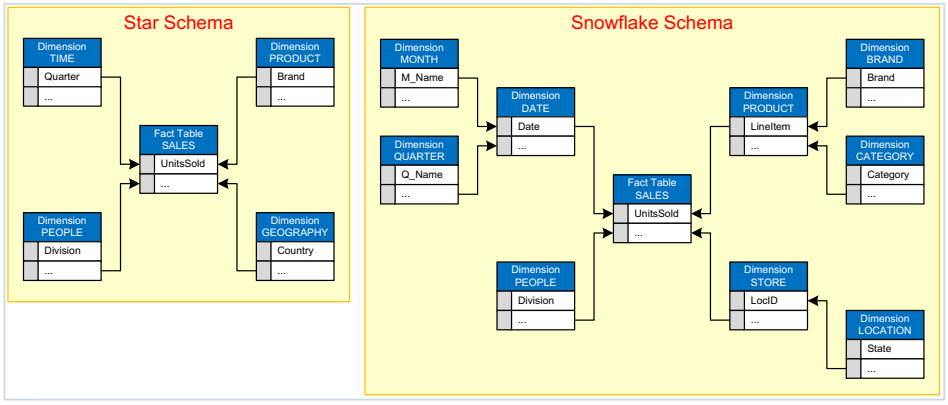
Multidimensionality

- 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)

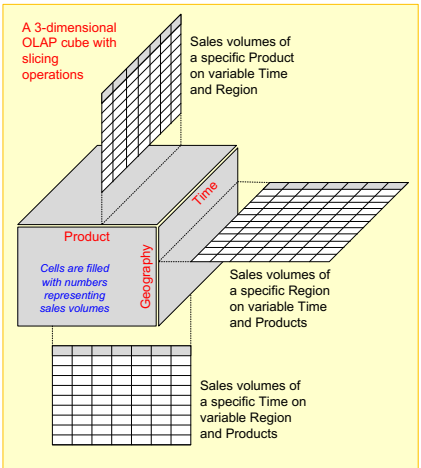
المعايير	OLTP	OLAP
الغرض	للقيام بمهام العمل اليومية	لدعم صنع القرار وتقديم إجابات لاستفسارات الأعمال والإدارة
مصدر البيانات	قاعدة بيانات المعاملات (مخزن البيانات المعتاد يركز في المقام الأول على الكفاءة والاتساق)	مخزن البيانات أو سوق البيانات (يركز مخزن البيانات غير العادي على الدقة والاكتمال)
التقارير	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

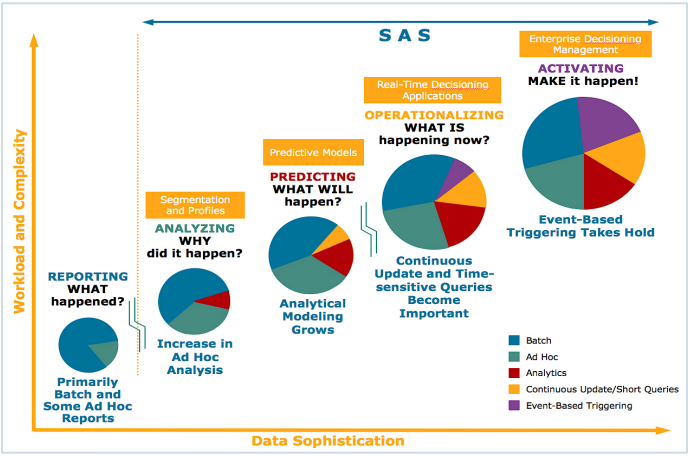
- قابلية التوسع
  - القضايا الرئيسية المتعلقة بقابلية التوسع:
    - كمية البيانات في المخزن
    - ما مدى سرعة نمو المخزن
    - عدد المستخدمين المتزامنين
    - تعقيد استعلامات المستخدم
  - تعني قابلية التوسع الجيد أن الاستعلامات ووظائف الوصول إلى البيانات الأخرى سوف تنمو بشكل خطي مع حجم المخزن

Real-Time/Active DW/BI

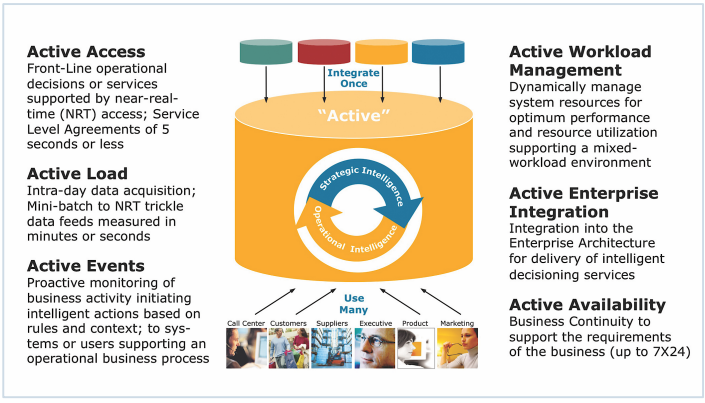
- Enabling real-time data updates for real-time analysis and real-time decision making is growing rapidly
  - Push vs. Pull (of data)
- Concerns about real-time BI
  - Not all data should be updated continuously
  - Mismatch of reports generated minutes apart
  - May be cost prohibitive
  - May also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

Business Reporting, Visual Analytics, and Business Performance Management

Business Reporting Definitions and Concepts

- Report = Information → Decision
- Report?
  - Any communication artifact prepared to convey specific information
- A report can fulfill many functions
  - To ensure proper departmental functioning
  - To provide information
  - To provide the results of an analysis
  - To persuade others to act
  - To create an organizational memory…

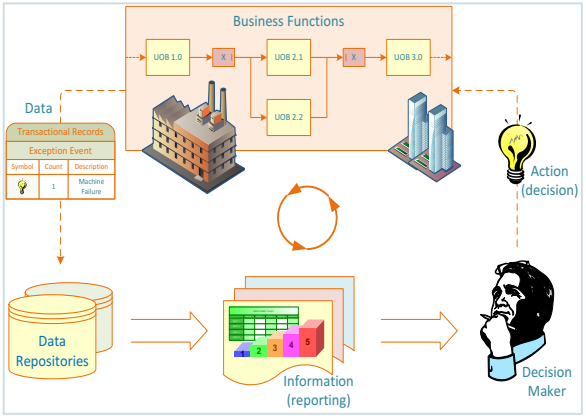
تقارير تعريفات الأعمال والمفاهيم

- التقرير = معلومات ← القرار
- التقرير؟
  - أي وسيلة او اداة للاتصال تم إعدادها لنقل معلومات محددة
- يمكن للتقرير أن يحقق العديد من الوظائف
  - لضمان حسن أداء الإدارات
  - لتوفير المعلومات
  - لتقديم نتائج التحليل
  - لإقناع الآخرين بالتصرف او العمل
  - لإنشاء ذاكرة تنظيمية …

What is a Business Report?

- A written document that contains information regarding business matters.
- **Purpose**: to improve managerial decisions
- **Source**: data from inside and outside the organization (via the use of ETL)
- **Format**: text + tables + graphs/charts
- **Distribution**: in-print, email, portal/intranet

Data acquisition → Information generation → Decision making → Process management



ما هو تقرير الاعمال

- وثيقة مكتوبة تحتوي على معلومات تتعلق بأمور الأعمال.
- **الغرض**: لتحسين القرارات الإدارية
- **المصدر**: بيانات من داخل وخارج المنظمة عبر استخدام (ETL)
- **التنسيق**: نص + جداول + رسوم بيانية / مخططات
- **التوزيع**: في الطباعة ، والبريد الإلكتروني ، وبوابة / إنترانت

الحصول على البيانات ← إنشاء المعلومات ← صنع القرار ← إدارة العمليات

Key to Any Successful Report

- Clarity …
- Brevity …
- Completeness …
- Correctness …
- Report types (in terms of content and format)
  - Informal – a single letter or a memo
  - Formal – 10-100 pages; cover + summary + text
  - Short report – periodic, informative, investigative

مفتاح ل اي تقرير ناجح

- وضوح …
- الاختصار …
- اكتمال …
- الصواب …
- أنواع التقارير (من حيث المحتوى والتنسيق)
  - غير رسمية - حرف واحد أو مذكرة
  - رسمي - 10-100 صفحة ؛ الغلاف + ملخص + النص
  - تقرير قصير - دوري ، إعلامي ، تحقيقي

Types of Business Reports

- Metric Management Reports
  - Help manage business performance through metrics (SLAs for externals; KPIs for internals) **(SLA = Service-level Agreement) , (KPL = Key Performance Indicator)**
  - Can be used as part of Six Sigma and/or TQM
- Dashboard-Type Reports
  - Graphical presentation of several performance indicators in a single page using dials/gauges
- Balanced Scorecard-Type Reports
  - Include financial, customer, business process, and learning & growth indicators

أنواع تقارير الاعمال

- تقارير إدارة المقياس
  - المساعدة في إدارة أداء الأعمال من خلال المقاييس (اتفاقيات مستوى الخدمة للعمليات الخارجية ؛ مؤشرات الأداء الرئيسية للداخلية)
  - يمكن استخدامها كجزء من Six Sigma و / أو TQM
- لوحة المعلومات – نوع التقارير
  - عرض بياني للعديد من مؤشرات الأداء في صفحة واحدة باستخدام الأوجه / المقاييس
- تقارير من نوع بطاقة الأداء المتوازن
  - وتشمل المالية ، والعملاء ، وعملية الأعمال ، ومؤشرات التعلم والنمو

Components of Business Reporting Systems

- Common characteristics
  - OLTP (online transaction processing)
    - ERP, POS, SCM, RFID, Sensors, Web, …
  - Data supply (volume, variety, velocity, …)
  - ETL
  - Data storage
  - Business logic
  - Publication medium
  - Assurance

مكونات نظم تقارير الأعمال

- الخصائص المشتركة
  - OLTP (معالجة المعاملات عبر الإنترنت)
    - ERP، POS، SCM، RFID، Sensors، Web،…
  - تزويد البيانات (الحجم ، التنوع ، السرعة ، ...)
  - ETL
  - مخزن البيانات
  - منطق الأعمال
  - وسط النشر
  - توكيد

Data and Information Visualization

"**The use of visual representations to explore, make sense of, and communicate data**."

- Data visualization vs. Information visualization
- Information = aggregation, summarization, and contextualization of data
- Related to information graphics, scientific visualization, and statistical graphics
- Often includes charts, graphs, illustrations, …

تصور البيانات و المعلومات

"استخدام التمثيلات المرئية لاستكشاف البيانات وإدراكها وتوصيلها."

- تصور البيانات Vs تصور المعلومات
- المعلومات = التجميع والتلخيص وسياق البيانات
- المتعلقة برسومات المعلومات ، والتصور العلمي ، والرسومات الإحصائية
- غالبًا ما يتضمن المخططات والرسوم البيانية والرسوم التوضيحية ...

A Brief History of Data Visualization

- Data visualization can date back to the second century AD
- Most developments have occurred in the last two and a half centuries
- Until recently it was not recognized as a discipline
- Today's most popular visual forms date back a few centuries

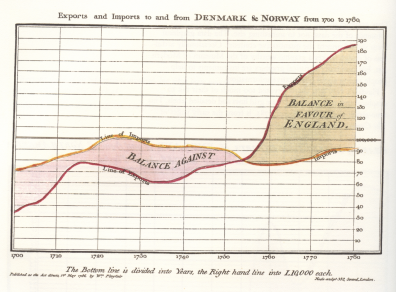
تاريخ موجز لتصور البيانات

- يمكن أن يرجع تاريخ عرض البيانات إلى القرن الثاني الميلادي
- حدثت معظم التطورات في القرنين ونصف القرن الماضي
- حتى وقت قريب ، لم يتم الاعتراف بها كنظام
- إن أكثر الأشكال المرئية شيوعًا اليوم تعود إلى بضعة قرون

The First Pie Chart Created by William Playfair in 1801

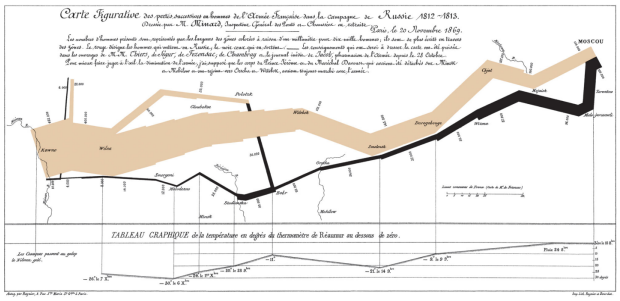
William Playfair is widely credited as the inventor of the modern chart, having created the first line and pie charts.

يُنسب الفضل إلى ويليام بلايفير على نطاق واسع كمخترع الرسم البياني الحديث ، بعد أن أنشأ الخط الأول والمخططات الدائرية.



Decimation of Napoleon's Army During the 1812 Russian Campaign

Arguably the most popular multi-dimensional chart → يمكن القول إن المخطط متعدد الأبعاد الأكثر شعبية



A Brief History of Data Visualization

- 1900s –
  - more formal attitude toward visualization
  - focus on color, value scales, and labeling
  - Publication of the book Semiologie Graphique
- 2000s –
  - Emergence of Internet as the medium for information visualization → raising visual literacy
  - Incorporate interaction, animation, 3D graphics-rendering, virtual worlds, real-time data feed
- 2010s and beyond – ?

تاريخ موجز لتصور البيانات

- 1900 -
  - موقف او سلوك أكثر رسمية تجاه التصور
  - التركيز على اللون ومقاييس القيمة ووضع العلامات
  - نشر كتاب Semiologie Graphique
- 2000 s-
  - ظهور الإنترنت كوسيلة لتصور المعلومات مما يرفع مستوى المعرفة البصرية
  - دمج التفاعل ، الرسوم المتحركة ، تقديم الرسومات ثلاثية الأبعاد ، العوالم الافتراضية ، تغذية البيانات في الوقت الحقيقي
- 2010s وما بعدها؟

Different Types of Charts and Graphs → أنواع مختلفة من المخططات و الرسوم البيانية

- Which one to use? Where and when?

Basic Charts and Graphs	Specialized Charts and Graphs
○ Line Chart ○ Bar Chart ○ Pie Chart ○ Scatter Plot ○ Bubble Chart	○ Histogram ○ Gantt Chart ○ PERT Chart ○ Geographic Map ○ Bullet Graph ○ Heat Map / Tree Map ○ Highlight Table

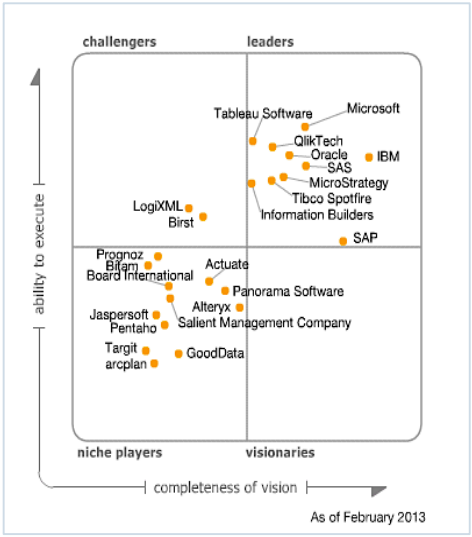
The Emergence of Data Visualization And Visual Analytics

- Magic Quadrant for Business Intelligence and Analytics Platforms (Source: Gartner.com)
  - Many data visualization companies are in the 4th quadrant
  - There is a move toward visualization

نشأة تصور البيانات و التحليلات البصرية

- Magic Quadrant تكون لمنصات التحليلات و ذكاء الاعمال
  - الكثير من شركات تصور البيانات تكون بالربع الرابع
  - هناك تحرك نحو التصور

- Emergence of new companies → نشأة شركات جديدة
  - Tableau, Spotfire, QlikView, …
- Increased focus by the big players → زيادة التركيز من قبل الممثلين الكبار
  - MicroStrategy improved Visual Insight
  - SAP launched Visual Intelligence
  - SAS launched Visual Analytics
  - Microsoft bolstered PowerPivot with Power View
  - IBM launched Cognos Insight
  - Oracle acquired Endeca



Visual Analytics

- A recently coined term
  - Information visualization + predictive analytics
- Information visualization
  - Descriptive, backward focused
  - "what happened" "what is happening"
- Predictive analytics
  - Predictive, future focused
  - "what will happen" "why will it happen"
- There is a strong move toward **visual analytics**

التحليلات البصرية او المرئية

- مصطلح صاغ مؤخرا
  - تصور المعلومات + التحليلات التنبؤية
- تصور المعلومات
  - وصفي ، التركيز العكسي
  - "ماذا حدث" "ما الذي يحدث"
- التحليلات التنبؤية
  - تنبؤية ، التركز في المستقبل
  - "ماذا سيحدث" "لماذا يحدث ذلك"
- هناك تحرك قوي نحو **التحليلات المرئية**

Visual Analytics by SAS Institute

- SAS Visual Analytics Architecture
  - Big data + In memory + Massively parallel processing + ..

هيكلة SAS للتحليلات البصرية
بيانات كبيرة + في الذاكرة + معالجة متوازية بشكل كبير + ..

❖ Data Builder ○ Join data from multiple sources ○ Create calculated and derived columns ○ Load data	❖ Administrator ○ Monitor SAS LASR ○ Analytic Server ○ Load/unload data ○ Manage security	❖ Explorer ○ Perform ad hoc analysis and data discovery ○ Apply advanced analytics	❖ Designer ○ Create dashboard style reports for web or mobile	❖ Mobile BI ○ Native iso and android applications that deliver interactive reports
❖ منشئ البيانات ○ ربط البيانات من مصادر متعددة ○ إنشاء أعمدة مدروسة ومستمدة ○ تحميل البيانات	❖ المدير ○ مراقب SAS LASR Analytic Server ○ تحميل / إلغاء تحميل البيانات ○ إدارة الأمن	❖ المستكشف ○ إجراء تحليل مخصص واكتشاف البيانات ○ تطبيق التحليلات المتقدمة	❖ المصمم ○ إنشاء تقارير نمط لوحة المعلومات للويب أو الهاتف المحمول	❖ Mobile BI ○ تطبيقات ios و android الأصلية التي تقدم تقارير تفاعلية

Performance Dashboards

- Performance dashboards are commonly used in BPM software suites and BI platforms
- Dashboards provide visual displays of important information that is consolidated and arranged on a single screen so that information can be digested at a single glance and easily drilled in and further explored

أداء لوحات المعلومات

- يتم استخدام لوحات بيانات الأداء بشكل شائع في مجموعات برامج BPM وأنظمة BI
- توفر لوحات المعلومات عروض مرئية لمعلومات مهمة يتم تجميعها وترتيبها على شاشة واحدة حتى يمكن دمج المعلومات في لمحة واحدة وحفرها بسهولة واستكشافها بشكل أكبر.

- Dashboard design
  - The fundamental challenge of dashboard design is to display all the required information on a single screen, clearly and without distraction, in a manner that can be assimilated quickly
- Three layer of information
  - Monitoring
  - Analysis
  - Management

- تصميم لوحة المعلومات
  - يتمثل التحدي الأساسي لتصميم لوحة المعلومات في عرض جميع المعلومات المطلوبة على شاشة واحدة بوضوح ودون إلهاء ، بطريقة يمكن استيعابها بسرعة
- ثلاثة طبقة من المعلومات
  - مراقبة
  - تحليل
  - إدارة

- What to look for in a dashboard
  - Use of visual components to highlight data and exceptions that require action.
  - Transparent to the user, meaning that they require minimal training and are extremely easy to use
  - Combine data from a variety of systems into a single, summarized, unified view of the business
  - Enable drill-down or drill-through to underlying data sources or reports
  - Present a dynamic, real-world view with timely data
  - Require little coding to implement/deploy/maintain

- ما الذي تبحث عنه في لوحة المعلومات
  - استخدام المكونات المرئية لتسليط الضوء على البيانات والاستثناءات التي تتطلب اتخاذ إجراء.
  - صريحه للمستخدم ، مما يعني أنه يتطلب الحد الأدنى من التدريب وهي سهلة الاستخدام للغاية
  - دمج البيانات من مجموعة متنوعة من الأنظمة في عرض واحد وموجز وموحد للأعمال
  - تمكين الانتقال إلى الأسفل أو الانتقال إلى مصادر أو تقارير البيانات الأساسية
  - تقديم عرض ديناميكي للعالم الحقيقي مع بيانات في الوقت المناسب
  - تتطلب القليل من الترميز لتنفيذ / نشر / صيانة

Best Practices in Dashboard Design

- Benchmark KPIs with Industry Standards
- Wrap the Metrics with Contextual Metadata
- Validate the Design by a Usability Specialist
- Prioritize and Rank Alerts and Exceptions
- Enrich Dashboard with Business-User Comments
- Present Information in Three Different Levels
- Pick the Right Visual Constructs
- Provide for Guided Analytics

أفضل الممارسات في تصميم لوحة المعلومات

- معيار KPIs مع معايير الصناعة
- لف المقاييس بالبيانات الوصفية السياقية
- التحقق من صحة التصميم بواسطة أخصائي قابلية الاستخدام
- تحديد الأولويات وترتيب التنبيهات والاستثناءات
- إثراء لوحة المعلومات مع تعليقات مستخدم الأعمال
- المعلومات الحالية في ثلاثة مستويات مختلفة
- اختر البناء البصري او المرئي الصحيح
- توفير التحليلات الإرشادية

Business Performance Management (BPM)

- **Business Performance Management (BPM) is…**
  - A real-time system that alerts managers to potential opportunities, impending problems and threats, and then empowers them to react through models and collaboration.
- Also called corporate performance management (CPM by Gartner Group), enterprise performance management (EPM by Oracle), strategic enterprise management (SEM by SAP)

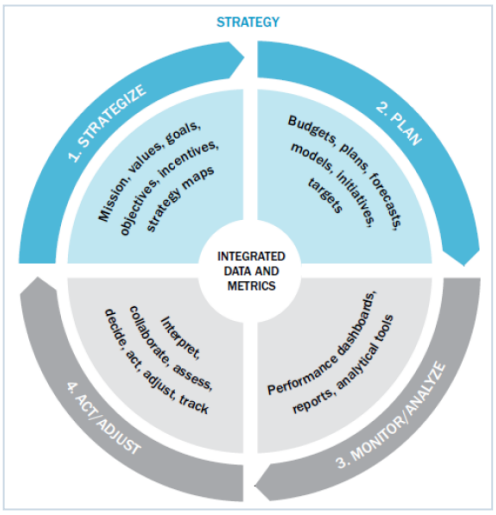
إدارة أداء الأعمال (BPM)

- **إدارة أداء الأعمال (BPM) هي** ...
  - نظام في الوقت الحقيقي يقوم بتنبيه المديرين إلى الفرص المحتملة والمشكلات والتهديدات الوشيكة ، ثم يمكّنهم من التفاعل من خلال النماذج والتعاون.
- وتسمى أيضًا إدارة أداء الشركات (CPM by Gartner Group) ، وإدارة أداء المؤسسات (EPM by Oracle)، وإدارة المشاريع الاستراتيجية (SEM by SAP)

- BPM refers to the business processes, methodologies, metrics, and technologies used by enterprises to measure, monitor, and manage business performance.
- BPM encompasses three key components
  - A set of integrated, closed-loop management and analytic processes, supported by technology …
  - Tools for businesses to define strategic goals and then measure/manage performance against them
  - Methods and tools for monitoring key performance indicators (KPIs), linked to organizational strategy

- يشير BPM إلى عمليات الأعمال والمنهجيات والمقاييس والتقنيات المستخدمة من قبل الشركات لقياس أداء الأعمال ومراقبته وإدارته.
- يشمل BPM ثلاثة مكونات رئيسية
  - مجموعة من الإدارة المتكاملة والحلقة المغلقة وعمليات التحليل ، مدعومة بالتكنولوجيا ...
  - أدوات للشركات لتحديد الأهداف الاستراتيجية ثم قياس / إدارة الأداء ضدهم
  - أساليب وأدوات رصد مؤشرات الأداء الرئيسية (KPIs)، المرتبطة بالاستراتيجية التنظيمية

A Closed-Loop Process to Optimize Business Performance



- Process Steps → خطوات العملية
  1. Strategize → وضع استراتيجية
  2. Plan → خطة
  3. Monitor/analyze → مراقبة / تحليل
  4. Act/adjust → التصرف / التعديل

Each with its own process steps → كل واحد مع خطوات العملية الخاصة فيه

Strategize: Where Do We Want to Go?

- Strategic planning
  - **Common tasks for the strategic planning process:**
    1. Conduct a current situation analysis
    2. Determine the planning horizon
    3. Conduct an environment scan
    4. Identify critical success factors
    5. Complete a gap analysis
    6. Create a strategic vision
    7. Develop a business strategy
    8. Identify strategic objectives and goals

وضع استراتيجية: أين نريد الذهاب؟

- تخطيط استراتيجي
  - **المهام الشائعة لعملية التخطيط الاستراتيجي:**
    1. إجراء تحليل الوضع الحالي
    2. تحديد أفق التخطيط
    3. إجراء مسح او فحص البيئة
    4. تحديد عوامل النجاح الحاسمة
    5. إكمال تحليل الفجوة
    6. خلق رؤية الاستراتيجية
    7. تطوير استراتيجية العمل
    8. تحديد غاية الاستراتيجية والاهداف

Plan: How Do We Get There?

- **Operational planning**
  - Operational plan: plan that translates an organization's strategic objectives and goals into a set of well-defined tactics and initiatives, resources requirements, and expected results for some future time period (usually a year).
- Operational planning can be
  - Tactic-centric (operationally focused)
  - Budget-centric plan (financially focused)

الخطة: كيف نصل إلى هناك؟

- **التخطيط العملي**
  - الخطة العملية او التشغيلية : الخطة التي تترجم الغايات والأهداف الاستراتيجية لمنظمة ما إلى مجموعة من التكتيكات والمبادرات المحددة جيدا ، ومتطلبات الموارد ، والنتائج المتوقعة لبعض الوقت في المستقبل (عادة في السنة).
- التخطيط التشغيلي يمكن أن يكون
  - تركز على التكتيك (يركز على الناحية التشغيلية)
  - خطة تتمحور حول الميزانية (تركز ماليًا)

Monitor/Analyze: How Are We Doing?

- A comprehensive framework for monitoring performance should address two key issues:

يجب أن يتناول إطار شامل لرصد الأداء قضيتين رئيسيتين:

  - What to monitor? → ماذا تراقب؟
    - Critical success factors → عوامل النجاح الحاسمة
    - Strategic goals and targets → أهداف وغايات الاستراتيجية
    - …
  - How to monitor? → كيف تراقب

Act and Adjust: What Do We Need to Do Differently?

- Success (or mere survival) depends on new projects: creating new products, entering new markets, acquiring new customers (or businesses), or streamlining some process.
- Many new projects and ventures fail!
- What is the chance of failure?
  - 60% of Hollywood movies fail
  - 70% of large IT projects fail, …

التصرف والتعديل: ما الذي يتعين علينا القيام به بشكل مختلف؟

- يعتمد النجاح (أو مجرد البقاء) على المشاريع الجديدة: خلق منتجات جديدة ، دخول أسواق جديدة ، اكتساب عملاء جدد (أو شركات) ، أو تبسيط بعض العمليات.
- العديد من المشاريع الجديدة والمغامرات تفشل!
- ما هي فرصة الفشل؟
  - 60 ٪ من أفلام هوليوود تفشل
  - 70 ٪ من مشاريع تكنولوجيا المعلومات الكبرى تفشل ، ...

Performance Measurement

- **Performance measurement system**

  A system that assists managers in tracking the implementations of business strategy by comparing actual results against strategic goals and objectives
  - Comprises systematic comparative methods that indicate progress (or lack thereof) against goals

مقياس الأداء

- **نظام قياس الأداء**

نظام يساعد المديرين في تتبع تنفيذ استراتيجية العمل من خلال مقارنة النتائج الفعلية مقابل الأهداف وغايات الاستراتيجية
  - يتضمن أساليب مقارنة منتظمة تشير إلى التقدم (أو عدمه) مقابل الأهداف

KPIs and Operational Metrics

- **Key performance indicator (KPI)**

  A KPI represents a strategic objective and metrics that measure performance against a goal
- Distinguishing features of KPIs
  - Strategy
  - Targets
  - Ranges
  - Encodings
  - Time frames
  - Benchmarks

مؤشرات الأداء الرئيسية والقياسات التشغيلية

- **مؤشر الأداء الرئيسي (KPI)**

يمثل مؤشر الأداء الرئيسي هدفًا استراتيجيًا ومقاييس لقياس الأداء مقابل هدف
- ميزات من KPIs
  - استراتيجية
  - أهداف
  - نطاقات
  - ترميزات
  - الاطارات الزمنية
  - المعايير

Performance Measurement

- **Key performance indicator (KBI) → مؤشر الأداء الرئيسي**

Outcome KPIs → نتيجة (lagging indicators e.g., revenues) (مؤشرات متأخرة مثل الإيرادات)

vs.

Driver KPIs → مشغل (leading indicators e.g., sales leads) (المؤشرات الرائدة ، على سبيل المثال ، المبيعات)

- Operational areas covered by driver KPIs → المناطق التشغيلية التي يغطيها مشغل KPIs
  - Customer performance → أداء العميل
  - Service performance → أداء الخدمة
  - Sales operations → عمليات المبيعات
  - Sales plan/forecast → خطة المبيعات / التوقعات

Performance Measurement System

- **Balanced Scorecard (BSC)**
  - A performance measurement and management methodology that helps translate an organization's financial, customer, internal process, and learning and growth objectives and targets into a set of actionable initiatives

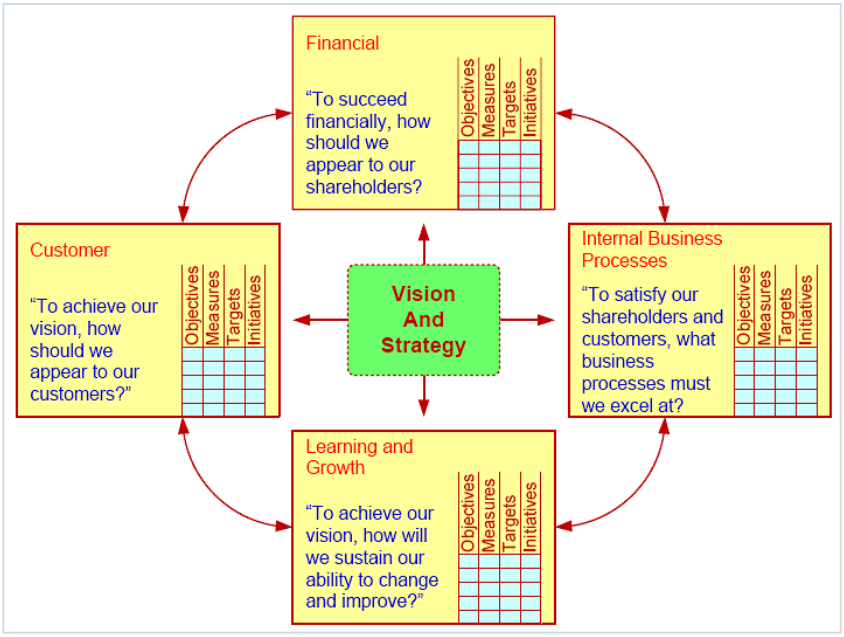
"The Balanced Scorecard: Measures That Drive Performance"

نظام قياس الأداء

بطاقة الأداء المتوازن (BSC)

- منهجية لقياس الأداء والإدارة تساعد في ترجمة أهداف المؤسسة المالية والعملاء والعمليات الداخلية وأهداف التعلم والنمو إلى مجموعة من المبادرات القابلة للتنفيذ

"بطاقة الأداء المتوازن: التدابير التي تدفع الأداء"



Six Sigma as a Performance Measurement System

- Six Sigma
  - A performance management methodology aimed at reducing the number of defects in a business process to as close to zero defects per million opportunities (DPMO) as possible
  - منهجية لإدارة الأداء تهدف إلى تقليل عدد العيوب في العمليات التجارية إلى ما يقرب من العيوب صفر لكل مليون فرصة (DPMO) قدر الإمكان
- The DMAIC performance model
  - A closed-loop business improvement model that encompasses the steps of **defining**, **measuring**, **analyzing**, **improving**, and **controlling** a process
  - نموذج تحسين الأعمال في حلقة مغلقة يشمل خطوات تحديد وقياس وتحليل وتحسين ومراقبة عملية
- Lean Six Sigma
  - Lean manufacturing / lean production
  - Lean production versus six sigma?

Comparison of Balanced Scorecard and Six Sigma → مقارنة بين بطاقة الأداء المتوازن و Six Sigma

TABLE 4.1 Comparison of Balanced Scorecard and Six Sigma

Balanced Scorecard	Six Sigma
Strategic management system	Performance measurement system
Relates to the longer-term view of the business	Provides snapshot of business's performance and identifies measures that drive performance toward profitability
Designed to develop balanced set of measures	Designed to identify a set of measurements that impact profitability
Identifies measurements around vision and values	Establishes accountability for leadership for wellness and profitability
Critical management processes are to clarify vision/strategy, communicate, plan, set targets, align strategic initiatives, and enhance feedback	Includes all business processes—management and operational
Balances customer and internal operations without a clearly defined leadership role	Balances management and employees' roles; balances costs and revenue of heavy processes

Six Sigma	بطاقة الأداء المتوازن
❖ **نظام قياس الأداء**	❖ **نظام الإدارة الاستراتيجية**
❖ يقدم لقطة عن أداء الأعمال ويحدد المقاييس التي تدفع الأداء نحو الربحية	❖ يربط عرض طويل الأجل للأعمال
❖ مصممة لتحديد مجموعة من القياسات التي تؤثر على الربحية	❖ مصممة لتطوير مجموعة متوازنة من المقاييس
❖ يحدد المساءلة عن القيادة من أجل الصحة والربحية	❖ يحدد القياسات حول الرؤية والقيم
❖ يشمل جميع عمليات إدارة الأعمال والتشغيلية	❖ إن عمليات الإدارة الحاسمة هي توضيح الرؤية / الاستراتيجية ، والتواصل ، والتخطيط ، وتحديد الأهداف ، ومواءمة المبادرات الاستراتيجية ، وتعزيز التعليقات
❖ أرصدة الإدارة وأدوار الموظفين ؛ أرصدة تكلفة وإيرادات العمليات الثقيلة	❖ أرصدة العملاء والعمليات الداخلية دون دور قيادي واضح المعالم

Data Mining

Data Mining Concepts/Definitions , Why Data Mining?

- More intense competition at the global scale.
- Recognition of the value in data sources.
- Availability of quality data on customers, vendors, transactions, Web, etc.
- Consolidation and integration of data repositories into data warehouses.
- The exponential increase in data processing and storage capabilities; and decrease in cost.
- Movement toward conversion of information resources into nonphysical form.

تعاريف/مفاهيم التنقيب في البيانات ، لماذا التنقيب في البيانات ؟

- منافسة أكثر حدة على المستوى العالمي.
- التعرف على القيمة في مصادر البيانات.
- توافر بيانات عالية الجودة على العملاء والبائعين والمعاملات والويب وما إلى ذلك.
- توحيد وتكامل مستودعات البيانات في مستودعات او مخازن البيانات.
- الزيادة الأسية في قدرات معالجة البيانات وتخزينها ؛ وانخفاض في التكلفة.
- التحرك نحو تحويل موارد المعلومات إلى صيغة غير مادية او غير ملموسه.

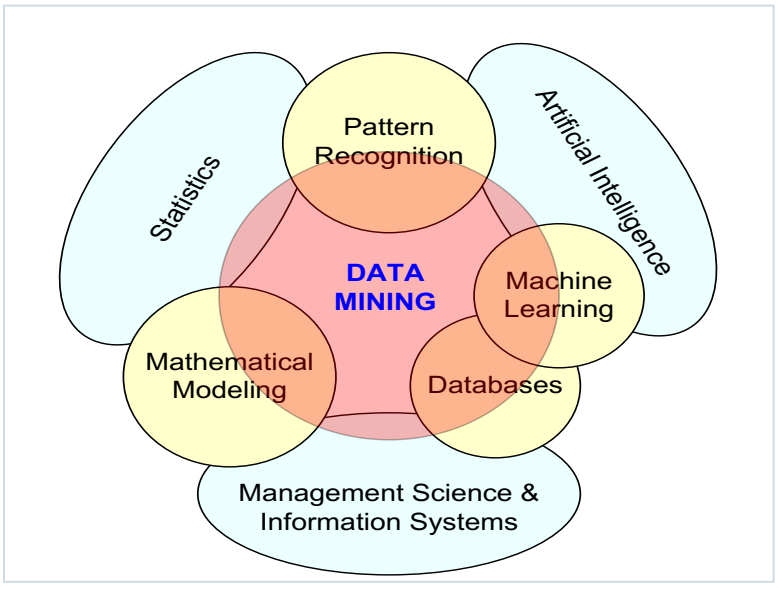
Definition of Data Mining

- The nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data stored in structured databases. - *Fayyad et al., (1996)*
- Keywords in this definition: Process, nontrivial, valid, novel, potentially useful, understandable.
- Data mining: a misnomer?
- Other names: knowledge extraction, pattern analysis, knowledge discovery, information harvesting, pattern searching, data dredging,…

تعريف التنقيب في البيانات

- عملية غير بديهية لتحديد أنماط صالحة ، جديدة ، قد تكون مفيدة ، ومفهومة في نهاية المطاف في البيانات المخزنة في قواعد البيانات المنظمة. - *Fayyad et al., (1996)*
- الكلمات الرئيسية في هذا التعريف: عملية ، غير بديهية ، صالحة ، جديدة ، يمكن أن تكون مفيدة ومفهومة.
- استخراج البيانات: تسمية خاطئة؟
- الأسماء الأخرى: استخراج المعرفة ، تحليل النمط ، اكتشاف المعرفة ، حصاد المعلومات ، البحث عن الأنماط ، تجريف البيانات ، ...

Data Mining is at the Intersection of Many Disciplines → التنقيب عن البيانات هو في تقاطع العديد من التخصصات



Data Mining Characteristics/Objectives

- Source of data for DM is often a consolidated data warehouse (not always!).
- DM environment is usually a client-server or a Web-based information systems architecture.
- Data is the most critical ingredient for DM which may include soft/unstructured data.
- The miner is often an end user
- Striking it rich requires creative thinking
- Data mining tools' capabilities and ease of use are essential (Web, Parallel processing, etc.)

خصائص و اهداف التنقيب في البيانات

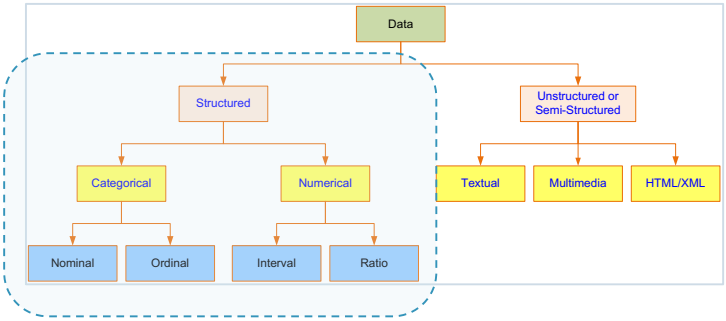
- مصدر البيانات لـ DM (Data Mining) هو مستودع بيانات مدمج (ليس دائمًا!).
- بيئة DM هي عادة خادم العميل أو بنية نظم المعلومات على شبكة الإنترنت.
- البيانات هي العنصر الأكثر أهمية ل DM والتي قد تتضمن بيانات هشة او ناقصة / غير منظمة.
- غالباً ما يكون عامل التنقيب مستخدمًا نهائيًا
- الحصول عليها غني يتطلب التفكير الإبداعي
- قدرة أدوات التنقيب عن البيانات وسهولة الاستخدام ضرورية (الويب ، المعالجة الموازية ، إلخ)

Data in Data Mining

- Data: a collection of facts usually obtained as the result of experiences, observations, or experiments.
- Data may consist of numbers, words, images, …
- Data: lowest level of abstraction (from which information and knowledge are derived).

البيانات في التنقيب عن البيانات

- البيانات: مجموعة من الحقائق التي يتم الحصول عليها عادة كنتيجة للتجارب أو الملاحظات أو التجارب.
- قد تتكون البيانات من أرقام ، كلمات ، صور ، ...
- البيانات: أدنى مستوى من التجريد (الذي تستمد منه المعلومات والمعارف).



What Does DM Do? How Does it Work?

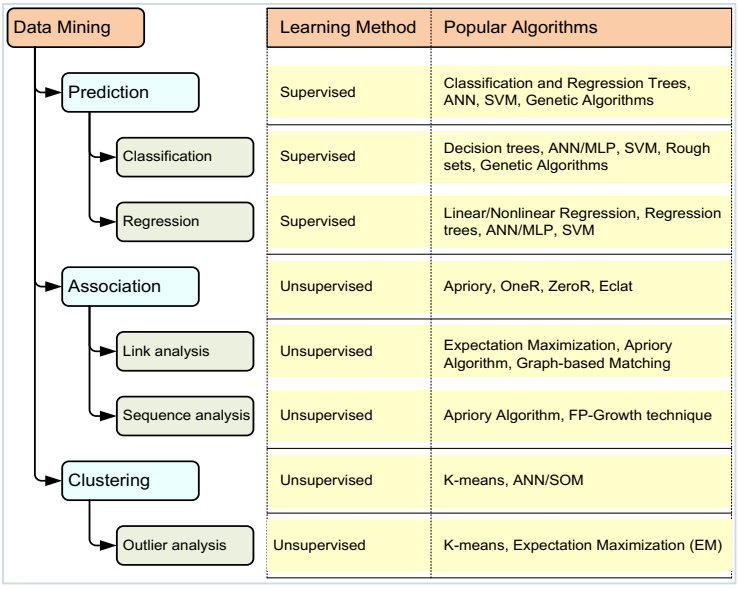
- DM extract patterns from data
  - Pattern? A mathematical (numeric and/or symbolic) relationship among data items
- Types of patterns
  - Association
  - Prediction
  - Cluster (segmentation)
  - Sequential (or time series) relationships

ماذا يعمل التنقيب عن البيانات و كيف يعمل ؟

- أنماط استخراج DM من البيانات
  - نمط؟ علاقة رياضية (رقمية و / أو رمزية) بين عناصر البيانات

- أنواع الأنماط
  - جمعية
  - تنبؤ
  - المجموعة (التجزئة)
  - العلاقات التسلسلية (أو السلاسل الزمنية)

A Taxonomy for Data Mining Tasks → تصنيف لمهام التنقيب عن البيانات



Learning Method	Popular Algorithms
Supervised	Classification and Regression Trees, ANN, SVM, Genetic Algorithms
Supervised	Decision trees, ANN/MLP, SVM, Rough sets, Genetic Algorithms
Supervised	Linear/Nonlinear Regression, Regression trees, ANN/MLP, SVM
Unsupervised	Apriory, OneR, ZeroR, Eclat
Unsupervised	Expectation Maximization, Apriory Algorithm, Graph-based Matching
Unsupervised	Apriory Algorithm, FP-Growth technique
Unsupervised	K-means, ANN/SOM
Unsupervised	K-means, Expectation Maximization (EM)

Data Mining Tasks (cont.)

- Time-series forecasting
  - Part of sequence or link analysis?
- Visualization
  - Another data mining task?
- Types of DM
  - Hypothesis-driven data mining
  - Discovery-driven data mining

مهام التنقيب عن البيانات

- التنبؤ بالسلسلة الزمنية
  - جزء من تسلسل أو تحليل الارتباط؟
- تصور
  - مهمة أخرى لتنقيب عن البيانات؟
- أنواع التنقيب عن البيانات
  - التنقيب عن البيانات منقاد بالفرضية
  - التنقيب عن البيانات منقاد بالاكتشاف

Data Mining Applications → تطبيقات التنقيب عن البيانات

- Customer Relationship Management
  - Maximize return on marketing campaigns
  - Improve customer retention (churn analysis)
  - Maximize customer value (cross-, up-selling)
  - Identify and treat most valued customers
- Banking & Other Financial
  - Automate the loan application process
  - Detecting fraudulent transactions
  - Maximize customer value (cross-, up-selling)
  - Optimizing cash reserves with forecasting

- إدارة علاقات العملاء
  - تعظيم العائد على الحملات التسويقية
  - تحسين احتفاظ العملاء (churn analysis)
  - تعظيم قيمة العملاء (cross-, up-selling)
  - تميز وتعامل مع معظم العملاء الكرام
- المصرفية وغيرها من المالية
  - أتمتة عملية طلب القرض
  - كشف المعاملات الاحتيالية
  - تعظيم قيمة العملاء (cross-, up-selling)
  - تحسين الاحتياطي النقدي مع التنبؤ

- Retailing and Logistics → البيع بالتجزئة و الخدمات اللوجستية (التخطيط و التنفيذ)
  - Optimize inventory levels at different locations → تحسين مستويات المخزون في مواقع مختلفة
  - Improve the store layout and sales promotions → تحسين تخطيط المتاجر و العروض الترويجية للمبيعات
  - Optimize logistics by predicting seasonal effects → تحسين اللوجستية عن طريق التنبؤ بالتأثيرات الموسمية
  - Minimize losses due to limited shelf life → تقليل الخسائر نظراً لمحدودية مدة الصلاحية
- Manufacturing and Maintenance → التصنيع و الصيانة
  - Predict/prevent machinery failures
  - Identify anomalies in production systems to optimize the use manufacturing capacity
  - Discover novel patterns to improve product quality
    - تنبأ / منع فشل الآلات
    - التعرف على الحالات الشاذة في أنظمة الإنتاج لتحسين استخدام القدرة التصنيعية
    - اكتشف أنماطًا جديدة لتحسين جودة المنتج

Data Mining Applications (cont.)

- Brokerage and Securities Trading
  - Predict changes on certain bond prices
  - Forecast the direction of stock fluctuations
  - Assess the effect of events on market movements
  - Identify and prevent fraudulent activities in trading
- Insurance
  - Forecast claim costs for better business planning
  - Determine optimal rate plans
  - Optimize marketing to specific customers
  - Identify and prevent fraudulent claim activities

- تداول الوساطة والأوراق المالية
  - تنبأ التغييرات على أسعار سندات معينة
  - توقع اتجاه تقلبات الأسهم
  - تقييم تأثير الأحداث على تحركات السوق
  - تحديد ومنع الأنشطة الاحتيالية في التداول

- تأمين
  - توقعات طلب التكاليف لتحسين تخطيط الأعمال
  - تحديد خطط المعدل الأمثل
  - تحسين التسويق لزبائن محددين و خاصيين
  - تحديد ومنع أنشطة المطالبة الاحتيالية

  - Computer hardware and software
  - Science and engineering
  - Government and defense
  - Homeland security and law enforcement
  - Travel industry
  - Healthcare
  - Medicine
  - Entertainment industry
  - Sports
  - Etc.

} Increasingly more popular application areas for data mining

  - أجهزة الكمبيوتر والبرمجيات
  - العلوم والهندسة
  - الحكومة والدفاع
  - الأمن الداخلي و تطبيق القانون
  - قطاع السفر
  - الرعاية الصحية
  - دواء
  - صناعه او مجال التسلية والترفيه
  - الأنشطة الرياضية

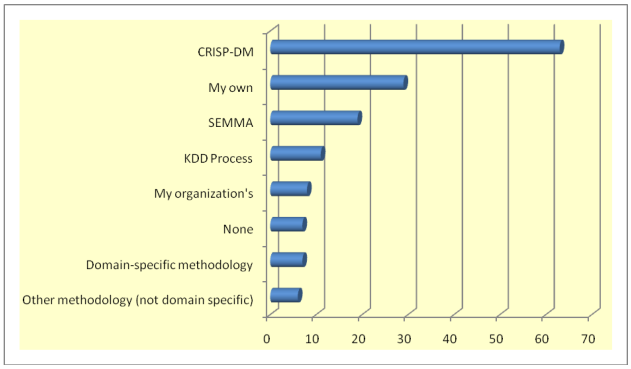
} مجالات التطبيق الأكثر انتشارًا لتنقيب عن البيانات

Data Mining Process → عملية التنقيب عن البيانات

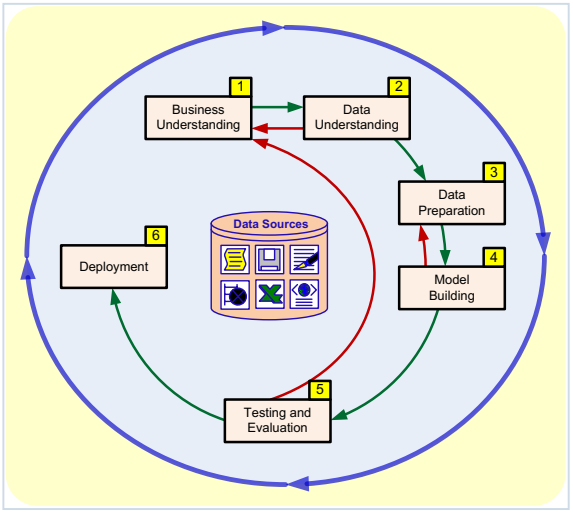
- A manifestation of best practices → مظهر من مظاهر أفضل الممارسات
- A systematic way to conduct DM projects → طريقة منظمة لإجراء مشاريع التنقيب عن البيانات
- Different groups has different versions → المجموعات المختلفة لديها إصدارات مختلفة
- Most common standard processes:
  - CRISP-DM (Cross-Industry Standard Process for Data Mining)
  - SEMMA (Sample, Explore, Modify, Model, and Assess)
  - KDD (Knowledge Discovery in Databases)

- العمليات القياسية الأكثر شيوعًا:
  - CRISP-DM (عملية قياسية عبر الصناعة لتنقيب عن البيانات)
  - SEMMA (عينه ، استكشاف ، تعديل ، نموذج ، وتقييم)
  - KDD (اكتشاف المعرفة في قواعد البيانات)

Data Mining Process



Data Mining Process: CRISP-DM



Step 1: Business Understanding
Step 2: Data Understanding
Step 3: Data Preparation (!)
Step 4: Model Building
Step 5: Testing and Evaluation
Step 6: Deployment

Accounts for ~85% of total project time

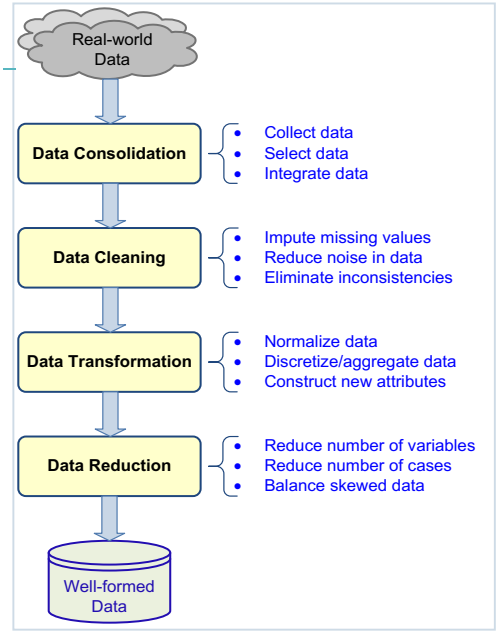
- The process is highly repetitive and experimental (DM: art versus science?)

الخطوة 1: فهم الأعمال
الخطوة 2: فهم البيانات
الخطوة 3: إعداد البيانات (!)
الخطوة 4: بناء النماذج
الخطوة 5: الاختبار والتقييم
الخطوة 6: النشر

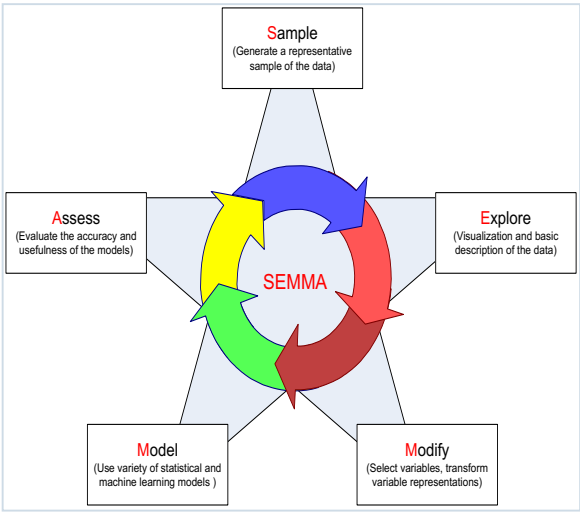
حسابات ~ 85٪ من إجمالي وقت المشروع

- العملية متكررة للغاية والتجريبية (DM: الفن مقابل العلم؟)

Data Preparation – A Critical DM Task



Data Mining Process: SEMMA



Data Mining Methods: Classification

- Most frequently used DM method
- Part of the machine-learning family
- Employ supervised learning
- Learn from past data, classify new data
- The output variable is categorical (nominal or ordinal) in nature
- Classification versus regression?
- Classification versus clustering?

طرق التنقيب عن البيانات : التصنيف

- طريقة DM الأكثر استخداما
- جزء من عائلة التعلم الآلي
- استخدام التعليم تحت الإشراف
- التعلم من البيانات السابقة ، وتصنيف البيانات الجديدة
- المتغير الناتج هو فئوي (اسمي أو ترتيبي) في الطبيعة - "او بمعنى خاص بفئة"
- التصنيف مقابل الانحدار او التراجع؟
- التصنيف مقابل التجمع؟

Assessment Methods for Classification

- Predictive accuracy
  - Hit rate
- Speed
  - Model building; predicting
- Robustness
- Scalability
- Interpretability
  - Transparency, explainability

طرق تقييم التصنيف

- دقة تنبؤية
  - معدل إصابة (الهدف)
- سرعة
  - بناء نموذج؛ توقعي
- متانة او صلابة
- قابلية التوسع
- للتفسير
  - الشفافية ، شرح

Accuracy of Classification Models → دقة نماذج التصنيف

- In classification problems, the primary source for accuracy estimation is the **confusion matrix**
- في مشاكل التصنيف ، يكون المصدر الأساسي لتقدير الدقة هو مصفوفة الارتباك او الحيرة

		True Class: Positive	True Class: Negative
Predicted Class	Positive	True Positive Count (TP)	False Positive Count (FP)
Predicted Class	Negative	False Negative Count (FN)	True Negative Count (TN)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

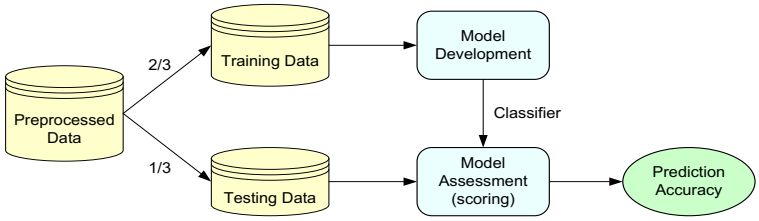
$$True\ Positive\ Rate = \frac{TP}{TP + FN}$$

$$True\ Negative\ Rate = \frac{TN}{TN + FP}$$

$$Precision = \frac{TP}{TP + FP} \qquad Recall = \frac{TP}{TP + FN}$$

Estimation Methodologies for Classification

- **Simple split** (or holdout or test sample estimation)
  - Split the data into 2 mutually exclusive sets training (~70%) and testing (30%)



  - For ANN, the data is split into three sub-sets (training [~60%], validation [~20%], testing [~20%])

منهجيات تقدير التصنيف

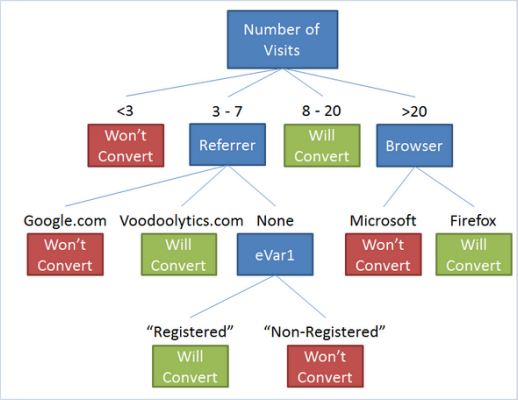
- انقسام بسيط (أو إجراء تقييم للعقد أو اختبار العينة)
  - قسّم البيانات إلى مجموعتين من الدورات التدريبية المتبادلة (~ 70٪) والاختبار (30٪)
  - بالنسبة إلى ANN ، يتم تقسيم البيانات إلى ثلاث مجموعات فرعية (تدريب [~ 60٪] ، التحقق من الصحة [~ 20٪] ، اختبار [~ 20٪])
- ***k*-Fold Cross** Validation (rotation estimation)
  - Split the data into *k* mutually exclusive subsets
  - Use each subset as testing while using the rest of the subsets as training
  - Repeat the experimentation for *k* times
  - Aggregate the test results for true estimation of prediction accuracy training
- Other estimation methodologies → منهجيات تقدير أخرى
  - **Leave-one-out, bootstrapping, jackknifing**
  - **Area under the ROC curve (**graph slide 33**)**

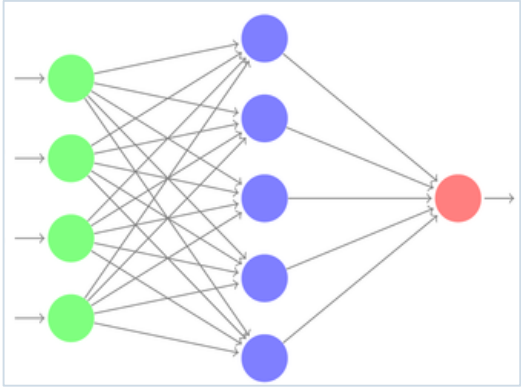
- فعالية k-Fold Cross (تقدير الدوران)
  - قسّم البيانات إلى *k* مجموعات فرعية حصرية متبادلة
  - استخدم كل مجموعة فرعية كاختبار أثناء استخدام بقية المجموعات الفرعية كتدريب
  - كرر تجربة k مرات
  - تجميع نتائج الاختبار لتقدير الصحيح لتدريب دقة التنبؤ

Classification Techniques

- Decision tree analysis
- Statistical analysis
- Neural networks
- Support vector machines
- Case-based reasoning
- Bayesian classifiers
- Genetic algorithms
- Rough sets

- تقنيات التصنيف
- تحليل شجرة القرار
- تحليل احصائي
- الشبكات العصبية
- دعم ناقلات الآلات
- الاستدلال المبني على حالة
- Bayesian classifiers
- الخوارزميات الجينية
- المجموعات قاسية





Decision Trees

- Employs the divide and conquer method
- Recursively divides a training set until each division consists of examples from one class

A general algorithm for decision tree building

1. Create a root node and assign all of the training data to it.
2. Select the best splitting attribute.
3. Add a branch to the root node for each value of the split. Split the data into mutually exclusive subsets along the lines of the specific split.
4. Repeat the steps 2 and 3 for each and every leaf node until the stopping criteria is reached.

شجرة القرارات

- تستخدم طريقة الانقسام والانتزاع
- يقسم مجموعة التدريب بشكل متكرر حتى يتكون كل قسم من أمثلة و من فصل واحد

خوارزمية عامة لبناء شجرة القرارات

1. إنشاء عقدة جذرية وتعيين جميع بيانات التدريب إليها.
2. حدد أفضل سمة التجزئة.
3. إضافة فرع إلى عقدة الجذر لكل قيمة من الانقسام. قسّم البيانات إلى مجموعات فرعية خاصة بالتبادل على طول خطوط التقسيم المحدد.
4. كرر الخطوتين 2 و 3 لكل عقدة كل ورقة حتى الوصول إلى معايير التوقف.

- DT algorithms mainly differ on
  1. Splitting criteria
     - Which variable, what value, etc.
  2. Stopping criteria
     - When to stop building the tree
  3. Pruning (generalization method)
     - Pre-pruning versus post-pruning
- Most popular DT algorithms include → الأكثر شعبية DT و تشمل خوارزميات
  - ID3, C4.5, C5; CART; CHAID; M5

- خوارزميات DT تختلف أساسا على
  1. معايير التقسيم
     - أي متغير وأي قيمة وما إلى ذلك
  2. وقف المعايير
     - متى تتوقف عن بناء الشجرة
  3. التقليم (طريقة التعميم)
     - قبل التقليم Vs ما بعد التقليم

- Alternative splitting criteria
  - Gini index determines the purity of a specific class as a result of a decision to branch along a particular attribute/value
    - Used in CART
  - Information gain uses entropy to measure the extent of uncertainty or randomness of a particular attribute/value split
    - Used in ID3, C4.5, C5
  - Chi-square statistics (used in CHAID)

- معايير تقسيم بديلة
  - يحدد مؤشر Gini نقاوة فئة معينة كنتيجة لقرار التفرع على طول سمة / قيمة معينة
    - تستخدم في CART
  - يستخدم اكتساب المعلومات entropy (هو مقاييس للعشوائية في النظام) لقياس مدى عدم اليقين أو العشوائية لتقسيم سمة / قيمة معينة
    - يستخدم في ID3, C4.5, C5

Cluster Analysis for Data Mining

- Used for automatic identification of natural groupings of things
- Part of the machine-learning family
- Employ unsupervised learning
- Learns the clusters of things from past data, then assigns new instances
- There is not an output variable
- Also known as segmentation

التحليل العنقودي (التجمعي) للتنقيب في البيانات

- تُستخدم للتعرف التلقائي على التجمعات الطبيعية للأشياء
- جزء من عائلة التعلم الآلي
- تستخدم التعلم بدون إشراف
- يتعلم مجموعات من الأشياء من البيانات السابقة ، ثم يعين حالات جديدة
- لا يوجد متغير النتيجة
- يُعرف أيضًا باسم التجزئة

- Clustering results may be used to
  - Identify natural groupings of customers
  - Identify rules for assigning new cases to classes for targeting/diagnostic purposes
  - Provide characterization, definition, labeling of populations
  - Decrease the size and complexity of problems for other data mining methods
  - Identify outliers in a specific domain (e.g., rare-event detection)

- نتائج المجموعات يمكن استخدامها ل
  - تحديد المجموعات الطبيعية للعملاء
  - حدد قواعد لتعيين حالات جديدة لفئات لأغراض الاستهداف / التشخيص
  - توفير توصيف وتعريف وتوسيم او تميز السكان
  - تقليل حجم وتعقيد المشاكل لطرق استخراج البيانات الأخرى
  - تحديد القيم المتطرفة في نطاق معين (على سبيل المثال ، اكتشاف الأحداث النادرة)

- Analysis methods
  - Statistical methods (including both hierarchical and nonhierarchical), such as *k*-means, *k*-modes, and so on.
  - Neural networks (adaptive resonance theory [ART], self-organizing map [SOM])
  - Fuzzy logic (e.g., fuzzy c-means algorithm)
  - Genetic algorithms

- طرق التحليل
  - الأساليب الإحصائية (بما في ذلك التسلسل الهرمي وغير الهرمي) ، مثل k-means و k-modes وما إلى ذلك.
  - الشبكات العصبية (نظرية الرنين التكيفي [ART]، خريطة التنظيم الذاتي [SOM])
  - المنطق الضبابي " يعني الغير واضح" (على سبيل المثال ، fuzzy c-means algorithm)
  - الخوارزميات الجينية

- How many clusters?
  - There is not a “truly optimal” way to calculate it
  - Heuristics are often used
- Most cluster analysis methods involve the use of a distance measure to calculate the closeness between pairs of items.
  - Euclidian versus Manhattan/Rectilinear distance

- كم عدد العناقيد؟
  - لا توجد طريقة "مثالية حقًا" لحسابها
  - وغالبا ما تستخدم الاستدلال
- تتضمن معظم طرق التحليل العنقودية استخدام مقياس المسافة لحساب التقارب بين أزواج العناصر.
  - Euclidian مقابل مسافة Manhattan/Rectilinear

Cluster Analysis for Data Mining (Cont.)

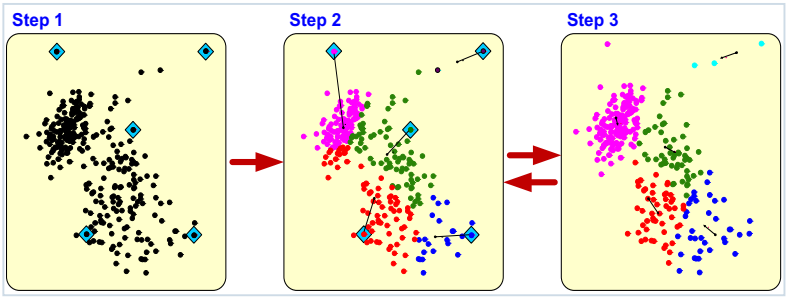
- *k*-Means Clustering Algorithm
  - ***k*** **:** pre-determined number of clusters
  - Algorithm (**Step 0:** determine value of *k*)

Step 1: Randomly generate *k* random points as initial cluster centers.
Step 2: Assign each point to the nearest cluster center.
Step 3: Re-compute the new cluster centers.
Repetition step: Repeat steps 3 and 4 until some convergence criterion is met (usually that the assignment of points to clusters becomes stable).

- تجميع خوارزمية k-Means
  - ***k:*** عدد محدد مسبقًا من العناقيد
  - الخوارزمية **الخطوة 0:** تحديد قيمة (k)

الخطوة 1: إنشاء نقاط عشوائية k كمراكز عنقودية أولية.
الخطوة 2: قم بتعيين كل نقطة إلى مركز العنقود الأقرب.
الخطوة 3: إعادة حساب مراكز العنقود الجديدة.
خطوة التكرار: كرر الخطوتين 3 و 4 حتى يتم استيفاء معيار تقارب (عادة ما يصبح تعيين النقاط إلى التجمعات او العناقيد مستقراً).

Cluster Analysis for Data Mining - *k*-Means Clustering Algorithm



Association Rule Mining

- A very popular DM method in business
- Finds interesting relationships (affinities) between variables (items or events)
- Part of machine learning family
- Employs unsupervised learning
- There is no output variable
- Also known as market basket analysis
- Often used as an example to describe DM to ordinary people, such as the famous "relationship between diapers and beers!"

تنظيم قاعدة التنقيب

- طريقة DM شعبية جدا في مجال الأعمال التجارية
- يجد العلاقات المثيرة للاهتمام (الانتماءات) بين المتغيرات (العناصر أو الأحداث)
- جزء من عائلة التعلم الآلي
- تستخدم التعلم بدون إشراف
- لا يوجد متغير النتيجة
- المعروف أيضا باسم تحليل سلة السوق
- غالبًا ما يستخدم كمثال لوصف DM للأشخاص العاديين .

- **Input:** the simple point-of-sale transaction data
- **Output:** Most frequent affinities among items
- Example: according to the transaction data…
  "Customer who bought a lap-top computer and a virus protection software, also bought extended service plan 70 percent of the time."
- How do you use such a pattern/knowledge?
  - Put the items next to each other
  - Promote the items as a package
  - Place items far apart from each other!

- **الإدخال:** بيانات معاملات نقطة البيع البسيطة
- **الإخراج:** أكثر الانتماءات المتكررة بين العناصر
- مثال: وفقًا لبيانات المعاملة ...

"العميل الذي اشترى جهاز كمبيوتر محمولًا وبرنامج حماية من الفيروسات ، اشترى أيضًا خطة خدمة موسعة بنسبة 70 بالمائة من الوقت".

- كيف تستخدم مثل هذا النمط / المعرفة؟
  - ضع العناصر بجانب بعضها البعض
  - تعزيز العناصر كحزمة
  - ضع العناصر بعيدًا عن بعضها البعض!

Association Rule Mining (Cont.)

- A representative applications of association rule mining include
  - **In business:** cross-marketing, cross-selling, store design, catalog design, e-commerce site design, optimization of online advertising, product pricing, and sales/promotion configuration
  - **In medicine:** relationships between symptoms and illnesses; diagnosis and patient characteristics and treatments (to be used in medical DSS); and genes and their functions (to be used in genomics projects)

- وتشمل التطبيقات التمثيلية للتنقيب قاعدة الارتباط
  - **في مجال الأعمال التجارية:** التسويق المتقاطع ، البيع المتقاطع ، تصميم المتجر ، تصميم الكتالوج ، تصميم موقع التجارة الإلكترونية ، تحسين الإعلان عبر الإنترنت ، تسعير المنتج ، وتكوين المبيعات / الترويج
  - **في الطب:** العلاقات بين الأعراض والأمراض. التشخيص وخصائص وعلاجات المريض (لاستخدامها في DSS الطبية) ؛ والجينات ووظائفها (لاستخدامها في مشاريع الجينوميات)

- Are all association rules interesting and useful? → هل كل قواعد الجمعيات مثيرة للاهتمام ومفيدة؟
  A Generic Rule: X ⇒ Y [S%, C%]
  X, Y: products and/or services → الخدمات او المنتجات
  X: Left-hand-side (LHS)
  Y: Right-hand-side (RHS)
  S: Support: how often X and Y go together → الدعم
  C: Confidence: how often Y go together with the X → الثقه
  Example: {Laptop Computer, Antivirus Software} ⇒ {Extended Service Plan} [30%, 70%]

- Algorithms are available for generating association rules
  - Apriori
  - Eclat
  - FP-Growth
  - + Derivatives and hybrids of the three
- The algorithms help identify the frequent item sets, which are, then converted to association rules

- الخوارزميات متاحة لتوليد قواعد الإتباط
  - Apriori
  - Eclat
  - FP-Growth
  - + مشتقه وهجين من الثلاثة
- تساعد الخوارزميات على تحديد مجموعات العناصر المتكررة ، والتي يتم تحويلها بعد ذلك إلى قواعد الارتباط

- **Apriori Algorithm**
  - Finds subsets that are common to at least a minimum number of the itemsets
  - Uses a bottom-up approach
    - frequent subsets are extended one item at a time (the size of frequent subsets increases from one-item subsets to two-item subsets, then three-item subsets, and so on), and
    - groups of candidates at each level are tested against the data for minimum support.

(*see the figure*) → --

- يجد مجموعات فرعية شائعة على الأقل العدد الاصغر لعدد العناصر
- يستخدم نهجًا من القاعدة إلى القمة
  - يتم توسيع مجموعات فرعية متكررة عنصر واحد في كل مرة (يزيد حجم المجموعات الفرعية بشكل متكرر من مجموعات فرعية عنصر واحد إلى مجموعات فرعية اثنين ، ثم مجموعات فرعية ثلاثة عناصر ، وهكذا) ، و
  - يتم اختبار مجموعات المرشحين على كل مستوى مقابل البيانات للحصول على الحد الأدنى من الدعم.

Raw Transaction Data

Transaction No	SKUs (Item No)
1	1, 2, 3, 4
1	2, 3, 4
1	2, 3
1	1, 2, 4
1	1, 2, 3, 4
1	2, 4

One-item Itemsets

Itemset (SKUs)	Support
1	3
2	6
3	4
4	5

Two-item Itemsets

Itemset (SKUs)	Support
1, 2	3
1, 3	2
1, 4	3
2, 3	4
2, 4	5
3, 4	3

Three-item Itemsets

Itemset (SKUs)	Support
1, 2, 4	3
2, 3, 4	3

Data Mining Software → هذي البرمجيات في الرسمه سلايد ٥٢

- Commercial → التجاري
  - IBM SPSS Modeler (formerly Clementine)
  - IBM - Intelligent Miner
  - … many more
  - SAS - Enterprise Miner
  - StatSoft – Statistica Data Miner
- Free and/or Open Source → مجاني او مفتوح المصدر
  - R
  - RapidMiner
  - Weka…

Data Mining Myths → اساطير التنقيب عن البيانات

- Data mining …
  - provides instant solutions/predictions
  - is not yet viable for business applications
  - requires a separate, dedicated database
  - can only be done by those with advanced degrees
  - is only for large firms that have lots of customer data
  - is another name for the good-old statistics

- التنقيب عن البيانات ....
  - يوفر حلول / توقعات فورية
  - ليست قابلة للتطبيق بعد لتطبيقات الأعمال
  - يتطلب قاعدة بيانات منفصلة ومخصصة
  - يمكن القيام به فقط من قبل أولئك الذين حصلوا على درجات علمية متقدمة
  - هو فقط للشركات الكبيرة التي لديها الكثير من بيانات العملاء
  - هو اسم آخر للإحصاءات القديمة

Common Data Mining Blunders → الأخطاء العامة للتنقيب عن البيانات

1. Selecting the wrong problem for data mining
2. Ignoring what your sponsor thinks data mining is and what it really can/cannot do
3. Not leaving insufficient time for data acquisition, selection and preparation
4. Looking only at aggregated results and not at individual records/predictions
5. Being sloppy about keeping track of the data mining procedure and results
6. …more in the book

1. اختيار مشكلة خاطئة لتعدين البيانات
2. تجاهل ما يعتقده الكفيل بالبيانات وماذا يمكن / لا يمكن فعله
3. عدم ترك الوقت غير الكافي لاكتساب البيانات ، والاختيار والتحضير
4. النظر فقط في النتائج المجمعة وليس في السجلات / التوقعات الفردية
5. كونها غير واضح بشأن تتبع إجراءات ونتائج استخراج البيانات

Techniques for Predictive Modeling

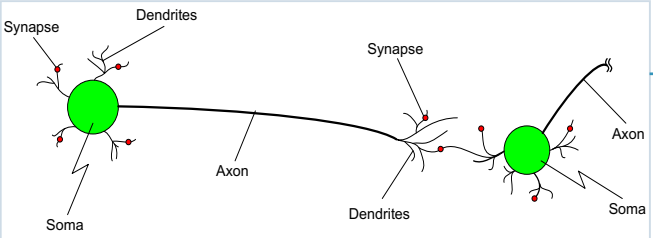
Neural Network Concepts

- **Neural networks (NN):** a brain metaphor for information processing
- Neural computing
- Artificial neural network (ANN)
- Many uses for ANN for
  - pattern recognition, forecasting, prediction, and classification
- Many application areas
  - finance, marketing, manufacturing, operations, information systems, and so on

مفهوم الشبكة العصبية

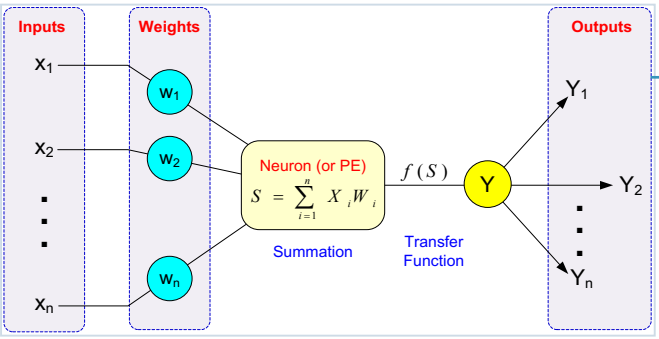
- **الشبكات العصبية (NN) :** استعارة دماغية لمعالجة المعلومات
- الحوسبة العصبية
- الشبكة العصبية الاصطناعية (ANN)
- العديد من الاستخدامات لـ ANN لـ
  - التعرف على الأنماط والتوقع والتنبؤ والتصنيف
- العديد من مجالات التطبيق
  - التمويل والتسويق والتصنيع والعمليات ونظم المعلومات ، وغيرها

Biological Neural Networks → الشبكات العصبية البيولوجية



- Two interconnected brain cells (neurons)
  اثنين من خلايا الدماغ المترابطة (الخلايا العصبية)

Processing Information in ANN → معالجة المعلومات في الشبكة العصبية الإصطناعية



- A single neuron (processing element – PE) with inputs and outputs
  خلية عصبيه وحده (عنصر المعالجة – PE) مع المدخلات و المخرجات

Biology Analogy → التناظر او التشابه علم الاحياء

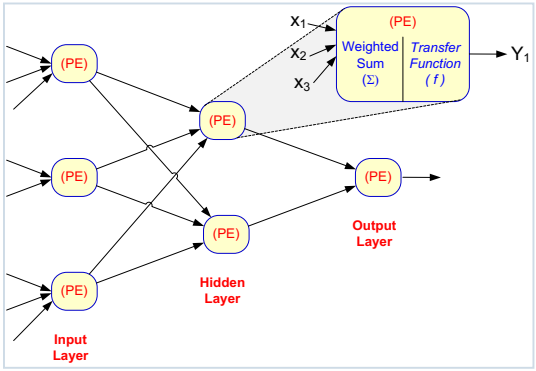
Biological – البيلوجي	Vs.	Artificial – الاصطناعي
Soma		Node
Dendrites - التشعبات		Input
Axon – محور عصبي		Output
Synapse – تشابك عصبي		Weight – له وزن
Slow – بطيء		Fast – سريع
Many neurons (10^9) – العديد من الخلايا العصبية		Few neurons (~100s) – عدد قليل من الخلايا العصبية

Elements of ANN

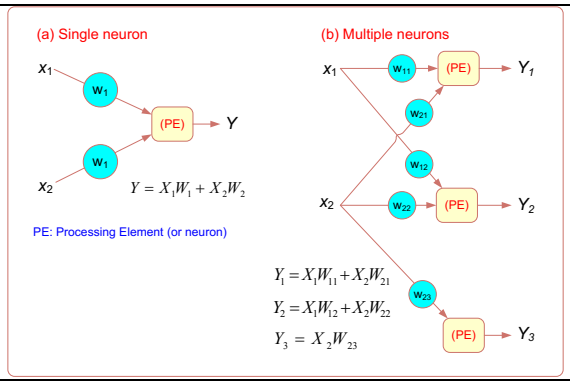
- Processing element (PE)
- Network architecture
  - Hidden layers
  - Parallel processing
- Network information processing
  - Inputs
  - Outputs
  - Connection weights
  - Summation function

عناصر ANN (الشبكة العصبية الاصطناعية)

- عنصر المعالجة (PE)
- بنية الشبكات
  - طبقات مخفية
  - المعالجة المتوازية
- معالجة معلومات الشبكة
  - المدخلات
  - المخرجات
  - أوزان التوصيل
  - دالة التلخيص



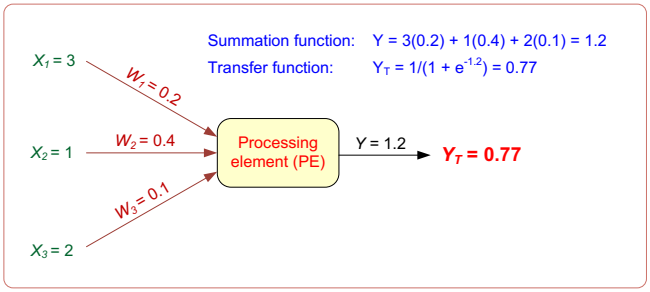
- Neural Network with One Hidden Layer
- الشبكة العصبية مع طبقة واحدة مخفية



- Summation Function for a Single Neuron (a), and Several Neurons (b)
- دالة التلخيص لخلية عصبية واحدة (a) ، و "العديد من الخلايا العصبية (ب)

Elements of ANN (Cont.)

- Transformation (Transfer) Function → دالة التحويل او النقل
  - Linear function → الدالة الخطية
  - Sigmoid (logical activation) function [0 1] → الدالة اللوجستية تكون للتنشيط المنطقي
  - Tangent Hyperbolic function [-1 1] → دالة المماس الزائدي او تسمى قوانين الدوال الزائدية



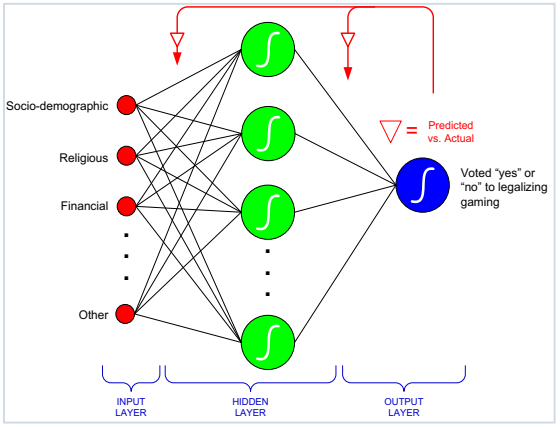
Neural Network Architectures

- Architecture of a neural network is driven by the task it is intended to address
  - Classification, regression, clustering, general optimization, association, ….
- **Most popular architecture:** Feedforward, multi-layered perceptron with backpropagation learning algorithm
  - Used for both classification and regression type problems
- **Others** – Recurrent, self-organizing feature maps, Hopfield networks, …

معمارية الشبكات العصبية

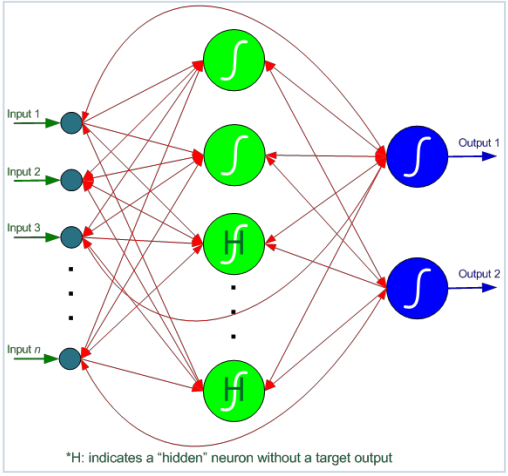
- البنية للشبكة العصبية تقاد بواسطة المهمة المقصودة بها العنوان
  - التصنيف ، التراجع ، التجميع ، التحسين العام ، الارتباط ، ...
- **البنيات الأكثر شعبية :** (هنا يذكر لي البنيات الي يتم استخدامها في طرق تعلم الشبكات العصبونية لتناقل المعلومات)
  - تستخدم لكل من مشاكل التصنيف والتراجع

Neural Network Architectures
(Feed-Forward Neural Networks)

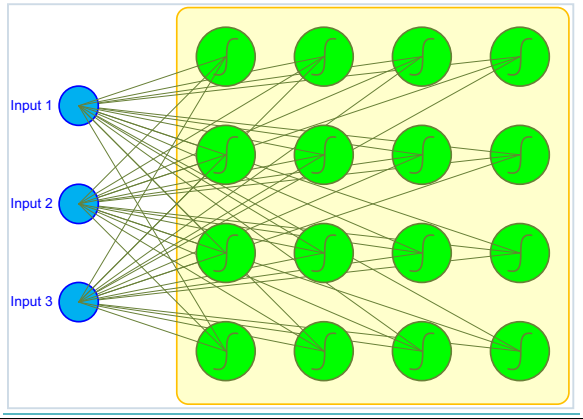


- Feed-forward MLP with 1 Hidden Layer

Neural Network Architectures
(Recurrent Neural Networks)



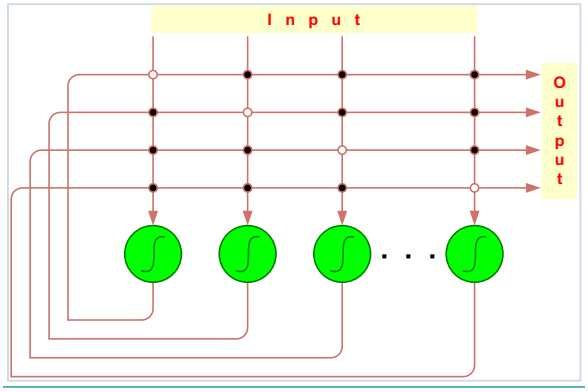
Other Popular ANN Paradigms
Self-Organizing Maps (SOM)



- First introduced by the Finnish Professor Teuvo Kohonen
- Applies to clustering type problems

- قدم لأول مرة من قبل الأستاذ الفنلندي Teuvo Kohonen
- ينطبق على نوع المشاكل العنقودية

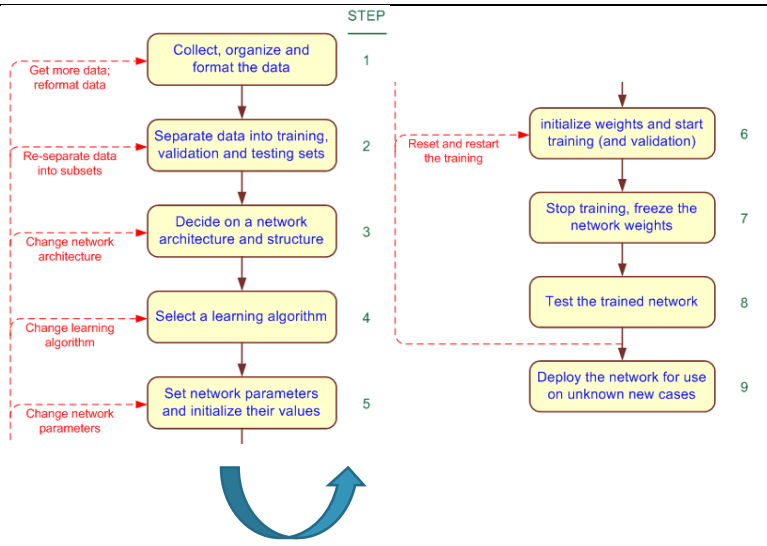
Other Popular ANN Paradigms
Hopfield Networks



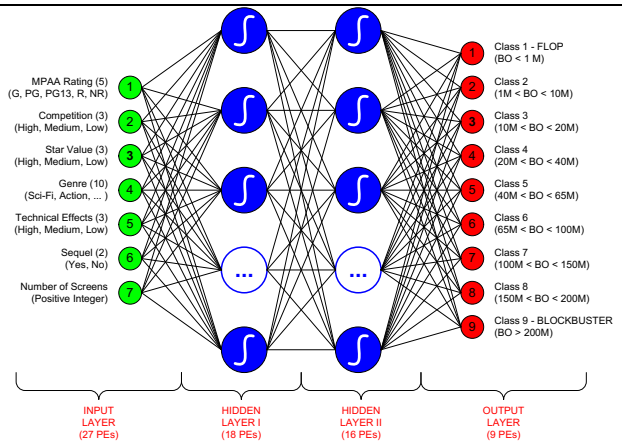
- First introduced by John Hopfield
- Highly interconnected neurons
- Applies to solving complex computational problems (e.g., optimization problems

- قدم لأول مرة من قبل John Hopfield
- الخلايا العصبية مترابطة للغاية
- تنطبق على حل المشكلات الحسابية المعقدة (مثل مشكلات التحسين)

Development Process of an ANN



An MLP ANN Structure for the Box-Office Prediction Problem



Testing a Trained ANN Model

- Data is split into three parts
  - Training (~60%)
  - Validation (~20%)
  - Testing (~20%)
- *k*-fold cross validation
  - Less bias
  - Time consuming

اختبار نموذج ANN المدرّب

- يتم تقسيم البيانات إلى ثلاثة أجزاء
  - التدريب (~ 60٪)
  - التحقق من الصحة (~ 20٪)
  - اختبار (~ 20٪)
- التحقق عبر k-fold
  - أقل تحيز
  - استهلاك الوقت

AN Learning Process A Supervised Learning Process

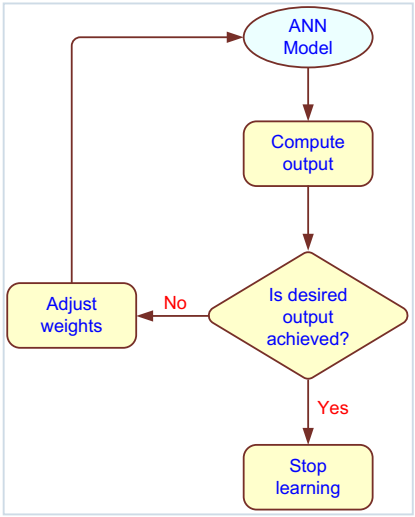
Three-step process:

1. Compute temporary outputs.
2. Compare outputs with desired targets.
3. Adjust the weights and repeat the process.

عملية التعلم AN عملية تتم بالإشراف

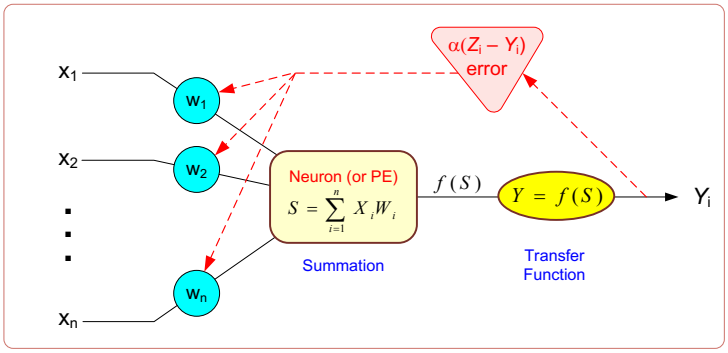
عملية من ثلاث خطوات:

1. حساب النواتج المؤقتة.
2. مقارنة المخرجات مع الأهداف المرجوة (المطلوبه).
3. ضبط الأوزان وتكرار العملية.



Backpropagation Learning

- Backpropagation of Error for a Single Neuron
- من الخطأ ان تكون لخلية عصبيه واحدة



- The learning algorithm procedure
  1. Initialize weights with random values and set other network parameters
  2. Read in the inputs and the desired outputs
  3. Compute the actual output (by working forward through the layers)
  4. Compute the error (difference between the actual and desired output)
  5. Change the weights by working backward through the hidden layers
  6. Repeat steps 2-5 until weights stabilize

- نهج خوارزمية التعلم
  1. قم بتهيئة الأوزان مع القيم العشوائية وقم بتعيين عوامل الشبكة الأخرى
  2. قراءة في المدخلات والمخرجات المطلوبة
  3. حساب الناتج الفعلي (من خلال العمل للأمام عبر الطبقات)
  4. حساب الخطأ (الفرق بين المخرجات الفعلية والمرغوبة)
  5. تغيير الأوزان من خلال العمل إلى الوراء من خلال الطبقات المخفية
  6. كرر الخطوات من 2 إلى 5 حتى تستقر الأوزان

Illuminating The Black Box Sensitivity Analysis on ANN

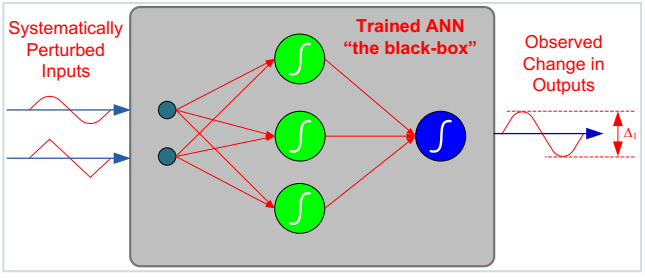
- A common criticism for ANN: The lack of transparency/explainability
- The black-box syndrome!
- Answer: sensitivity analysis
  - Conducted on a trained ANN
  - The inputs are perturbed while the relative change on the output is measured/recorded
  - Results illustrate the relative importance of input variables

إلقاء الضوء على الصندوق الأسود تحليل الحساسية على ANN

- انتقاد شائع لـ ANN: غياب الشفافية / التفسير
- متلازمة الصندوق الأسود!
- الجواب: تحليل الحساسية
  - أجريت على مدربين ANN
  - تشوش المدخلات بينما يتم قياس / تسجيل التغيير النسبي على المخرج
  - توضح النتائج أهمية النسبية لمتغيرات المدخلات

Sensitivity Analysis on ANN Models

- Sensitivity analysis reveals the most important injury severity factors in traffic accidents
- يكشف تحليل الحساسية عن أهم عوامل شدة الخطورة في حوادث المرور



Support Vector Machines (SVM)

- SVM are among the most popular machine-learning techniques.
- SVM belong to the family of generalized linear models… (capable of representing non-linear relationships in a linear fashion).
- SVM achieve a classification or regression decision based on the value of the linear combination of input features.
- Because of their architectural similarities, SVM are also closely associated with ANN.

الات ناقلات الدعم (SVM)

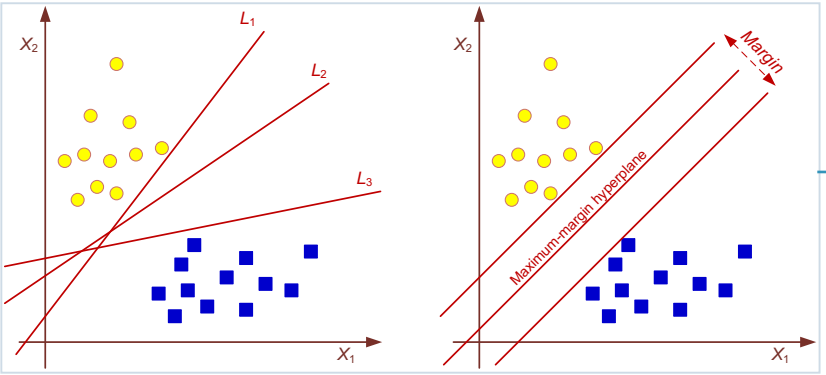
- SVM هي من بين تقنيات التعلم الآلي الأكثر شعبية.
- تنتمي SVM إلى عائلة النماذج الخطية المعممة ... (القادرة على تمثيل العلاقات غير الخطية بطريقة خطية).
- يحقق SVM قرار تصنيف أو تراجع يستند إلى قيمة التركيبة الخطية لميزات الإدخال.
- وبسبب التشابه المعماري او البنية ، ترتبط SVM أيضًا ارتباطًا وثيقًا بـ ANN.

- Goal of SVM: to generate mathematical functions that map input variables to desired outputs for classification or regression type prediction problems.
  - First, SVM uses nonlinear kernel functions to transform non-linear relationships among the variables into linearly separable feature spaces.
  - Then, the maximum-margin hyperplanes are constructed to optimally separate different classes from each other based on the training dataset.
- SVM has solid mathematical foundation!

- الهدف من SVM: لإنشاء دوال حسابية تقوم بتعيين متغيرات الإدخال إلى المخرجات المرغوبة من أجل تصنيف أو مشاكل التنبؤ بنوع التراجع.
  - أولاً ، يستخدم SVM دالة kernel غير الخطية لتحويل العلاقات غير الخطية بين المتغيرات إلى مسافات ميزة قابلة للفصل خطيًا.
  - بعد ذلك ، يتم إنشاء تعامد ذات حد أقصى للهامش لفصل الطبقات المختلفة عن بعضها البعض بشكلٍ مثالي استنادًا إلى مجموعة بيانات التدريب.
- SVM له أساس رياضي متين!

- A **hyperplane** is a geometric concept used to describe the separation surface between different classes of things.
  - In SVM, two parallel hyperplanes are constructed on each side of the separation space with the aim of maximizing the distance between them.
- **A kernel function** in SVM uses the kernel trick (a method for using a linear classifier algorithm to solve a nonlinear problem)
  - The most commonly used kernel function is the radial basis function (RBF).

- **و التعامد (الجبر خطي)** هو مفهوم هندسي يستخدم لوصف سطح الانفصال بين فئات مختلفة من الأشياء.
  - في SVM ، يتم إنشاء تعامديين فرعيين متوازيين على كل جانب من مساحة الفصل بهدف زيادة المسافة بينهما.
- تستخدم **دالة kernel** في SVM خدعة kernel (طريقة لاستخدام خوارزمية تصنيف خطي لحل مشكلة غير خطية)
  - دالة kernel الأكثر استخدامًا هي دالة الأساس الشعاعي (RBF)



- Many linear classifiers (hyperplanes) may separate the data
- قد يفصل العديد من المصنفات الخطية (hyperplanes) البيانات .

How Does an SVM Work?

- Following a machine-learning process, an SVM learns from the historic cases.
- The Process of Building SVM

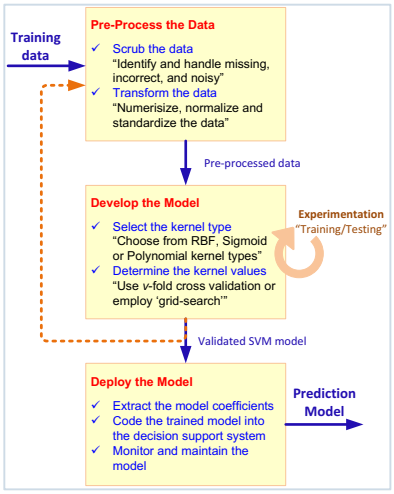
1. Preprocess the data
   - Scrub and transform the data.
2. Develop the model.
   - Select the kernel type (RBF is often a natural choice).
   - Determine the kernel parameters for the selected kernel type.
   - If the results are satisfactory, finalize the model; otherwise change the kernel type and/or kernel parameters to achieve the desired accuracy level.
3. Extract and deploy the model.

كيف تعمل SVM ؟

- بعد عملية تعلم الآلة ، يتعلم SVM من الحالات التاريخية.
- عملية بناء SVM

1. المعالجة المسبقة للبيانات
   - تنظيف وتحويل البيانات.
2. تطوير النموذج.
   - حدد نوع kernel (RBF غالبًا ما يكون اختيارًا طبيعيًا).
   - حدد عوامل kernel لنوع kernel المحدد.
   - إذا كانت النتائج مرضية ، قم بوضع الصيغة النهائية للنموذج ؛ خلاف ذلك تغيير نوع kernel و / أو عوامل kernel لتحقيق مستوى الدقة المطلوبة.
3. استخراج ونشر النموذج.

The Process of Building an SVM



SVM Applications

- SVMs are the most widely used kernel-learning algorithms for wide range of classification and regression problems
- SVMs represent the state-of-the-art by virtue of their excellent generalization performance, superior prediction power, ease of use, and rigorous theoretical foundation
- Most comparative studies show its superiority in both regression and classification type prediction problems.
- SVM versus ANN?

تطبيقات SVM

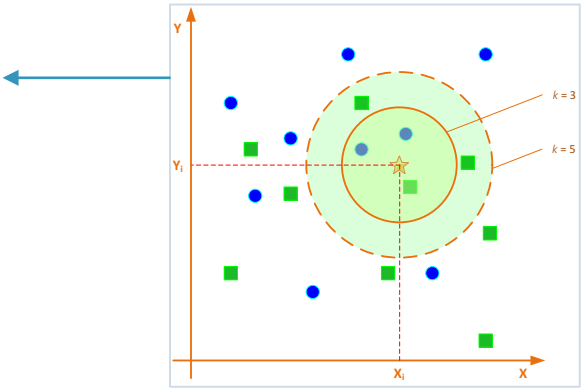
- SVMs هي أكثر خوارزميات تعلم kernel المستخدمة على نطاق واسع لمجموعة واسعة من مشاكل التصنيف والتراجع
- تمثل SVMs أحدث ما توصلت إليه التكنولوجيا بفضل أداءها الممتاز في التعميم ، وقوة التنبؤ الفائقة ، وسهولة الاستخدام ، والأساس النظري الصارم
- تظهر معظم الدراسات المقارنة تفوقها في كل من مشاكل التنبؤ بالتراجع والتصنيف.

k-Nearest Neighbor Method (k-NN)

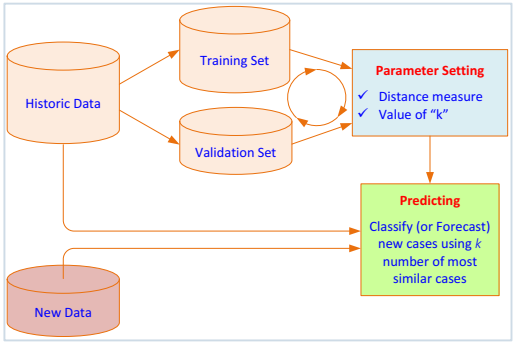
- ANNs and SVMs → time-demanding, computationally intensive iterative derivations
- *k*-NN is a simplistic and logical prediction method, that produces very competitive results
- *k*-NN is a prediction method for classification as well as regression types (similar to ANN & SVM)
- *k*-NN is a type of instance-based learning (or lazy learning) – most of the work takes place at the time of prediction (not at modeling)
- *k* : the number of neighbors used

- ANNs و SVMs ← التي تتطلب وقت ، اشتقاق متكرر مكثف حسابياً
- k-NN هي طريقة تنبؤ مبسطة ومنطقية ، تنتج نتائج تنافسية للغاية
- k-NN هي طريقة تنبؤ للتصنيف بالإضافة إلى أنواع التراجع المشابهة لـ (ANN & SVM)
- k-NN هو نوع من التعلم القائم على التثقيف (أو التعلم البطيء) - معظم العمل يحدث في وقت التنبؤ (وليس في النمذجة)
- k :عدد المتجاور المستخدم

- The answer depends on the value of *k*
- الجواب يعتمد على قيمة k



The Process of *k*-NN Method



k-NN Model Parameter

1. Similarity Measure: The Distance Metric → مقاييس التشابه : مقاييس المسافة

Minkowski distance

$$d(i,j) = \sqrt[q]{(|x_{i1} - x_{j1}|^q + |x_{i2} - x_{j2}|^q + \ldots + |x_{ip} - x_{jp}|^q)}$$

If $q = 1$, then d is called Manhattan distance

$$d(i,j) = \sqrt{|x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \ldots + |x_{ip} - x_{jp}|}$$

If $q = 2$, then d is called Euclidean distance

$$d(i,j) = \sqrt{(|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \ldots + |x_{ip} - x_{jp}|^2)}$$

   - Numeric versus nominal values? → ؟ القيم العددية مقابل القيم الأسمية

2. Number of Neighbors (the value of *k*)
   - The best value depends on the data
   - Larger values reduce the effect of noise but also make boundaries between classes less distinct
   - An "optimal" value can be found heuristically

- Cross Validation is often used to determine the best value for k and the distance measure

2 . عدد الجيران (هي قيمة k)
   - أفضل قيمة تعتمد على البيانات
   - القيم الأكبر تقلل من تأثير الضوضاء ولكنها أيضًا تجعل الحدود بين الطبقات أقل وضوحًا
   - يمكن العثور على قيمة "مثالية" بشكل تجريبياً
- وكثيرا ما يستخدم التحقق من صحة لتحديد أفضل قيمة ل k ومقياس المسافة