

---

# **Chương 1**

# **Giới thiệu**

---

# Nội dung

- ❑ Định nghĩa BI
- ❑ Công dụng của BI
- ❑ Khả năng của BI
- ❑ Lợi ích của BI
- ❑ Kiến trúc của BI
- ❑ Khuynh hướng công nghệ của BI

# Các định nghĩa Thông minh doanh nghiệp hay Trí tuệ doanh nghiệp - (BI)

- ❑ **Business intelligence (BI)** là thuật ngữ chung, chỉ *kỹ năng, qui trình, công nghệ, ứng dụng và thực hành* dùng để hỗ trợ việc đưa ra các quyết định dựa trên chứng cứ, trong các tổ chức, doanh nghiệp
- ❑ Theo nghĩa rộng hơn, hệ thống thông minh doanh nghiệp là tập hợp các tiếp cận để *thu thập, lưu trữ, phân tích và cung cấp cách truy cập dữ liệu*, nhằm giúp người dùng hiểu sâu sắc và đưa ra các quyết định tốt hơn, dựa trên các sự kiện nền tảng.
- ❑ Có thể hiểu BI là 'Giải pháp quản trị doanh nghiệp thông minh'

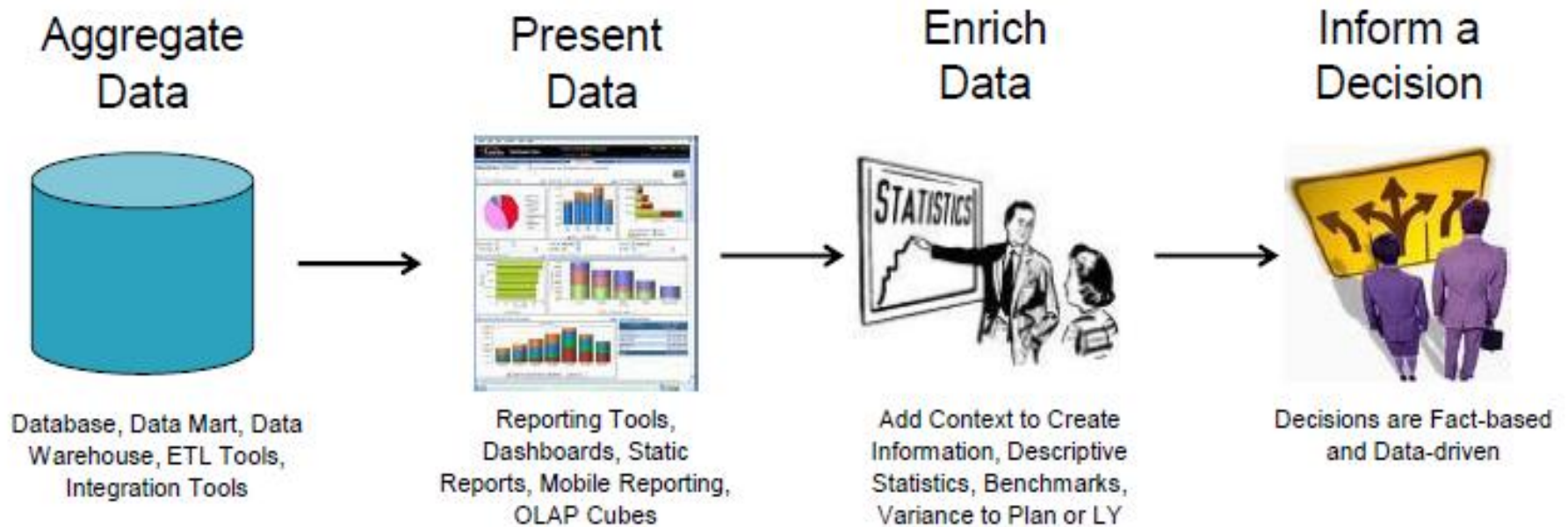
# **Các định nghĩa Thông minh doanh nghiệp hay Trí tuệ doanh nghiệp - (BI)**

- ❑ BI là tập các công nghệ và công cụ để chuyển đổi những dữ liệu thô thành những thông tin có nghĩa và có ích cho mục đích phân tích kinh doanh
- ❑ BI là các ứng dụng và công nghệ để chuyển dữ liệu doanh nghiệp thành hành động
- ❑ BI là công nghệ mới giúp doanh nghiệp hiểu biết về quá khứ và dự đoán tương lai
- ❑ BI trong tương lai gần sẽ tập trung vào các sự kiện thời gian thực và các dự báo nổi bật trong tương lai

# Thông minh doanh nghiệp (BI)

- ❑ **Phương pháp luận thông minh doanh nghiệp (BI)** rộng và liên quan đến nhiều lĩnh vực, trải dài qua nhiều phạm vi của ứng dụng. Nó liên quan đến:
  - Việc biểu diễn và tổ chức của quá trình ra quyết định
  - Phạm vi của lý thuyết quyết định: thu thập (collecting) và lưu trữ (storing) dữ liệu, giúp quá trình ra quyết định dễ dàng hơn.
  - Các công nghệ kho dữ liệu (**data warehousing**)
  - Các mô hình toán học dành cho tối ưu hóa và khai thác dữ liệu (**data mining**), các thao tác nghiên cứu và thống kê
  - Nhiều lĩnh vực ứng dụng như tiếp thị, kho vận (logistics), kế toán, điều khiển, tài chính, dịch vụ, hành chính công (public administration).

# Cơ sở của BI



## BI dùng để làm gì?

- ❑ Các tổ chức/doanh nghiệp dùng BI để hiểu sâu sắc hướng dữ liệu (data-driven) của mọi thứ liên quan đến hiệu quả kinh doanh của doanh nghiệp. BI dùng để hiểu và cải tiến hiệu năng thực hiện, cắt giảm chi phí, nhận biết cơ hội kinh doanh mới...
- ❑ Phân tích hành vi của khách hàng, các khuôn mẫu mua hàng (buying patterns), các khuynh hướng bán hàng (sales trends)...
- ❑ Đo lường, theo dõi, dự báo việc bán hàng và hiệu quả tài chính.

## BI dùng để làm gì?

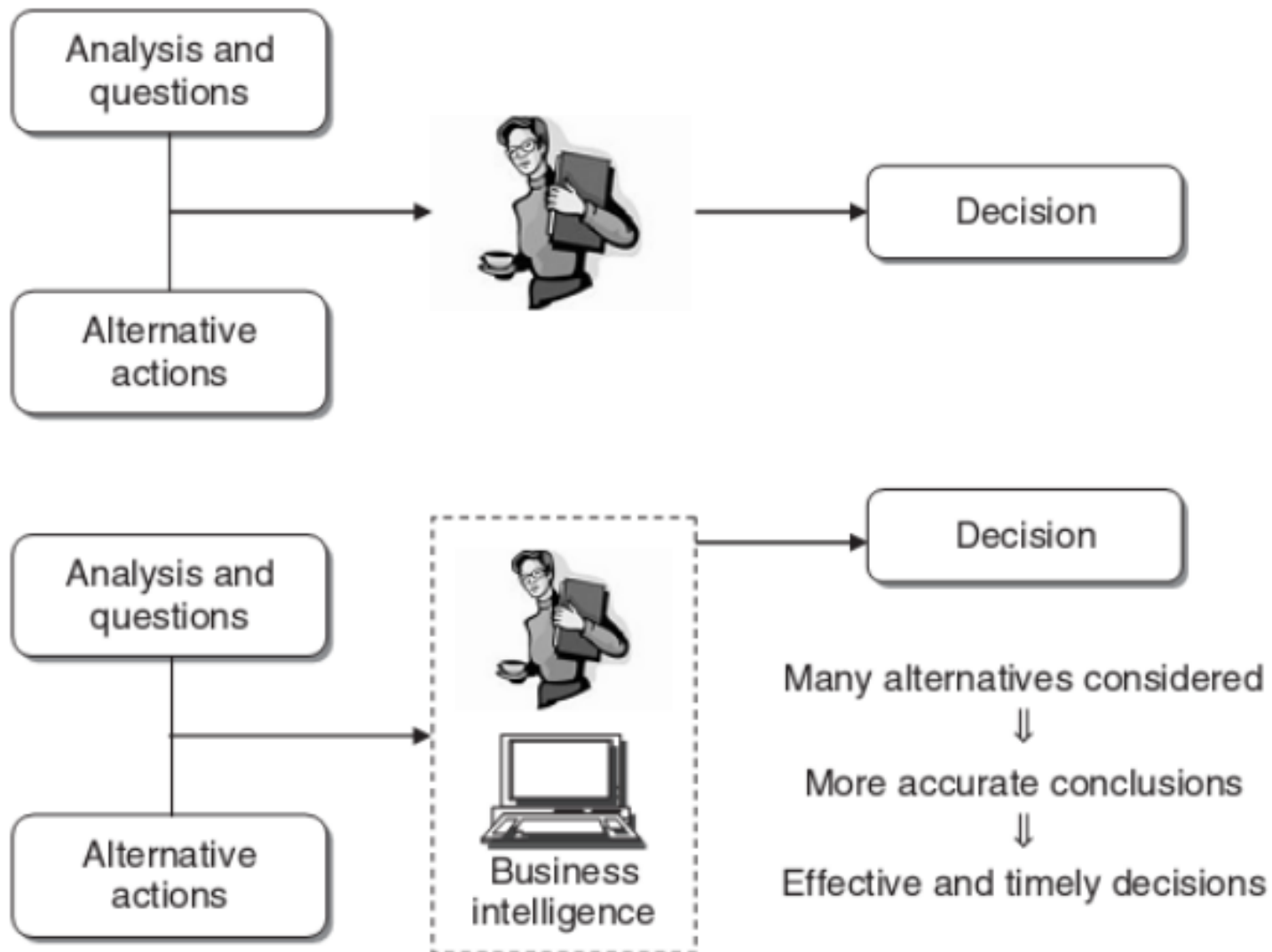
- ❑ Hoạch định và dự báo ngân sách và tài chính
- ❑ Theo dõi thực hiện các chiến dịch tiếp thị
- ❑ Tối ưu quá các quá trình và việc thực hiện hoạt động.
- ❑ Cải tiến việc giao nhận (delivery) và tính hiệu quả của chuỗi cung ứng (supply chain effectiveness)
- ❑ Phân tích Web và thương mại điện tử (e-commerce)
- ❑ Quản lý quan hệ khách hàng
- ❑ Phân tích rủi ro
- ❑ Phân tích hướng đến giá trị chiến lược (Strategic value driver analysis)



# Khả năng của BI

- ❑ Khả năng cung cấp **thông tin chính xác** khi cần, bao gồm cả góc nhìn thời gian thực của doanh nghiệp/ tổ chức
- ❑ Theo nghiên cứu của Thompson (2004):
  - Cho các báo cáo nhanh và chính xác hơn (81%)
  - Cải thiện việc ra quyết định (78%)
  - Cải thiện dịch vụ khách hàng (56%)
  - Tăng lợi nhuận (49%)

# So sánh (hệ thống có BI và không có BI)



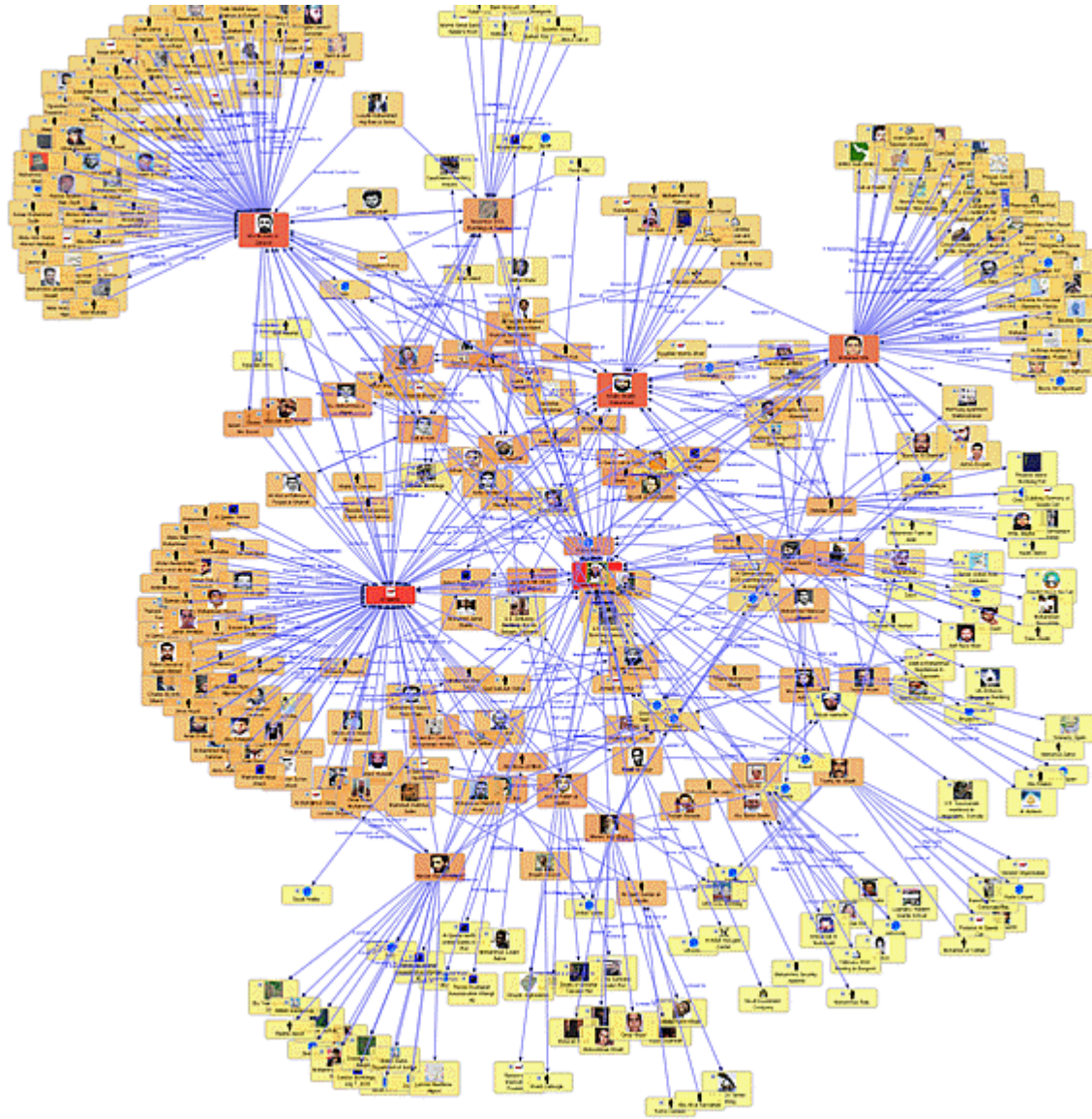
## Lợi ích của BI

- ❑ Cải tiến qui trình quản lý ( lên kế hoạch, kiểm soát, đo lường và/hoặc thay đổi kết quả nhằm tăng thu nhập, giảm chi phí)
- ❑ Cải tiến quá trình hoạt động (phát hiện lỗi, xử lý đơn hàng, mua bán... nhằm cho kết quả tăng thu nhập, giảm chi phí)
- ❑ Dự báo tương lai

## Bài toán

- ❑ Phân tích giỏ hàng (Market Basket Analytics)
- ❑ Phân tích văn bản (Text Analytics)
- ❑ Phân khúc/ Gom nhóm khách hàng (Customer Segmentation/Clustering)
- ❑ Phân hạng sản phẩm phù hợp (Tailored Product Assortments)
- ❑ Dự báo tồn kho (Inventory Forecasting )

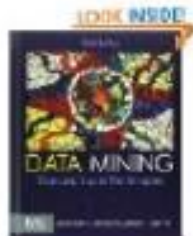
# Phân tích mạng XH (Social Network Analysis)



# Ví dụ về BI

- ❑ **Lọc cộng tác (Collaborative Filtering)** cố dự báo các sản phẩm khác mà khách hàng muốn mua, dựa trên giỏ hàng thực tế của khách và hành vi mua sắm của các khách hàng khác

## Customers Who Bought This Item Also Bought



Data Mining: Concepts and Techniques, Third...

› Jiawei Han

★★★★☆ 25

Hardcover

\$41.02



Data Science for Business: What you need...

› Foster Provost

★★★★☆ 88

Paperback

\$35.83



Business Intelligence Roadmap: The Complete...

Larissa T. Moss

★★★★☆ 26

Paperback

\$56.18



Big Data, Big Analytics: Emerging Business...

› Michael Minelli

★★★★☆ 18

Hardcover

\$35.45



# Ví dụ về BI



# Quyết định hiệu quả và quyết định đúng lúc

❑ *"Get the right information to the right people at the right time"*

❑ **Quyết định hiệu quả (Effective decisions):**

cho phép người ra quyết định dựa vào các thông tin và kiến thức đáng tin cậy để đưa ra các quyết định tốt và nghĩ ra các kế hoạch hành động cho phép đạt được mục tiêu theo cách thức hiệu quả nhất

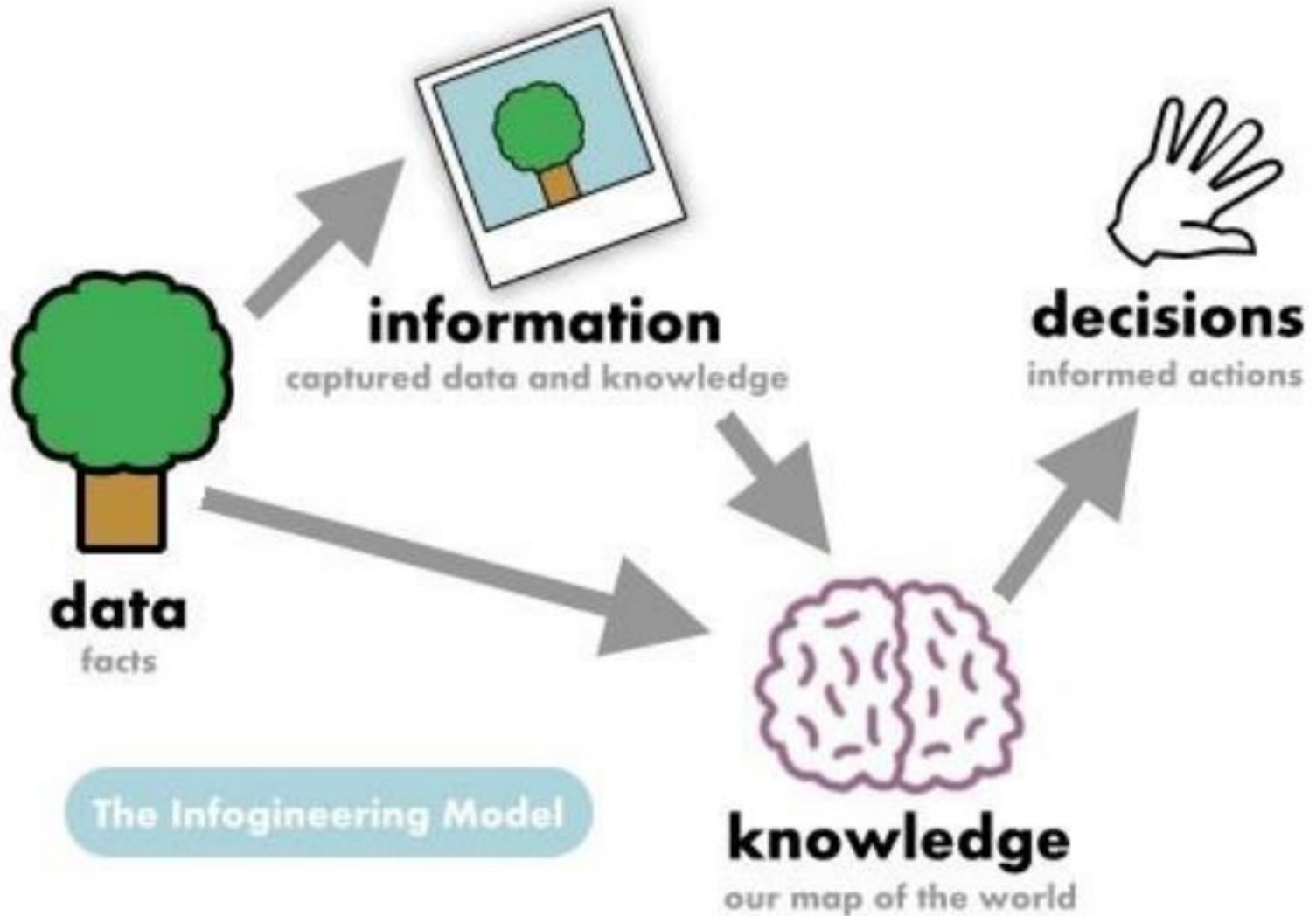
❑ **Quyết định đúng lúc (Timely decisions):**

doanh nghiệp làm việc trong môi trường đặc trưng bởi sự gia tăng mức độ cạnh tranh và tính năng động cao.

Hệ quả: khả năng phản ứng nhanh đối với hành động của các đối thủ cạnh tranh và đối với các điều kiện của thị trường mới là nhân tố quyết định cho sự thành công hoặc thậm chí sự sống còn của doanh nghiệp/công ty



# Dữ liệu, thông tin và tri thức



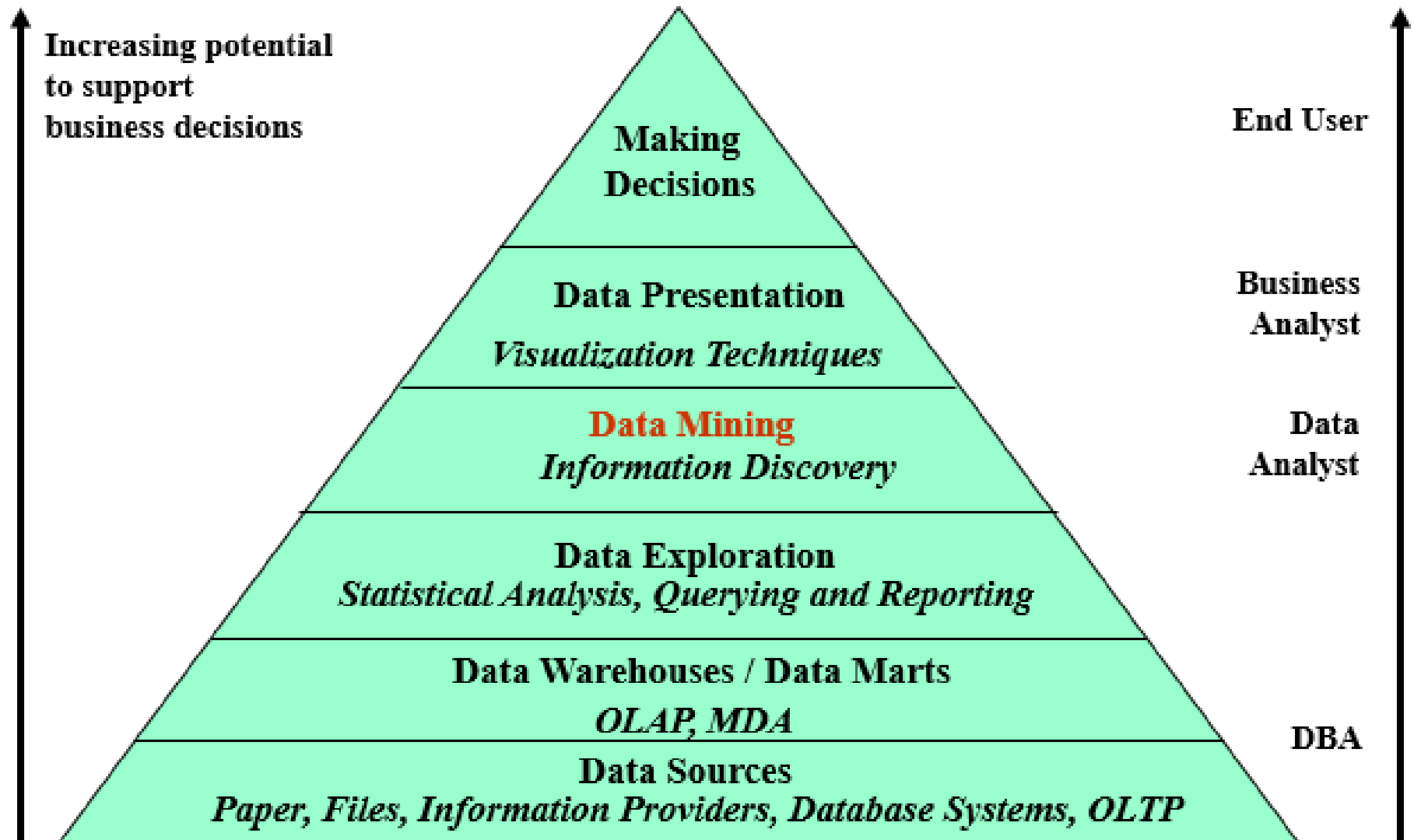
# Dữ liệu, thông tin và tri thức

- ❑ **Dữ liệu (Data):** thô, tồn tại ở dạng nào đó, tiện dụng hoặc không; bản thân dữ liệu không mang ý nghĩa nào
  - Tuy nhiên, ngay cả khi được thu thập và lưu trữ theo cách thức có hệ thống và có cấu trúc, dữ liệu cũng không thể trực tiếp dùng cho các mục tiêu ra quyết định.
- ❑ **Thông tin (Information):** là dữ liệu “có ý nghĩa” nhờ cách kết nối quan hệ. “Ý nghĩa” ở đây có thể hữu ích.
  - Trong máy tính, cơ sở dữ liệu quan hệ tạo ra thông tin từ dữ liệu được lưu trữ bên trong nó.

# Dữ liệu, thông tin và tri thức

- ❑ **Tri thức (Knowledge):** là tập hợp các thông tin có liên quan, dùng để ra quyết định và phát triển các hành động phù hợp.
  - Hoạt động cung cấp sự hỗ trợ về kiến thức cho công nhân tri thức qua việc tích hợp quá trình ra quyết định và cho phép dùng công nghệ thông tin, được coi là quản lý tri thức
- ❑ *Data: là các ký hiệu/biểu tượng*
- ❑ *Information: là data được xử lý để trở nên hữu ích, cung cấp câu trả lời cho câu hỏi "who", "what", "where", "when".*
- ❑ *Knowledge: là ứng dụng của data và information; trả lời câu hỏi "how"*

# Kiến trúc của BI



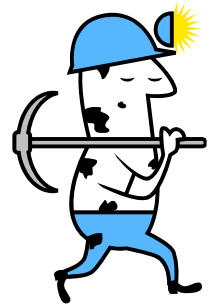
# Kiến trúc của BI

Gồm 6 thành phần chính:

- ❑ **Data sources:** gồm hầu hết các thành phần của dữ liệu thuộc về các hệ thống hoạt động, nhưng cũng có thể bao gồm các tài liệu phi cấu trúc, như email và các dữ liệu từ các nguồn cung cấp bên ngoài.
- ❑ **Data warehouses và data marts:**
  - Trích (extract), biến đổi (transform) và nạp (load) dữ liệu
  - Phân tích khai thác nhiều chiều (Multidimensional Exploratory Analysis)
- ❑ **Data exploration:** các công cụ để thực hiện các phân tích thông minh nhân tạo tự động, bao gồm các hệ thống truy vấn và báo cáo, cũng như các phương pháp thống kê

# Kiến trúc của BI

❑ **Data mining:** bao gồm các phương pháp luận thông minh nhân tạo tích cực, mà mục tiêu của nó là trích ra thông tin và kiến thức từ dữ liệu



– Các mô hình toán học dành cho nhận dạng mẫu (pattern recognition), machine learning và kỹ thuật khai phá dữ liệu

❑ **Data presentation và Making decision** (Biểu diễn dữ liệu và ra quyết định): cho phép xác định giải pháp tốt nhất trong tập các hành động lựa chọn, tập này thường khá bao quát và đôi khi vô hạn

⇒ Vấn đề cốt lõi trong hệ thống BI là kho dữ liệu (Data Warehouse) và khai phá dữ liệu (Data Mining)

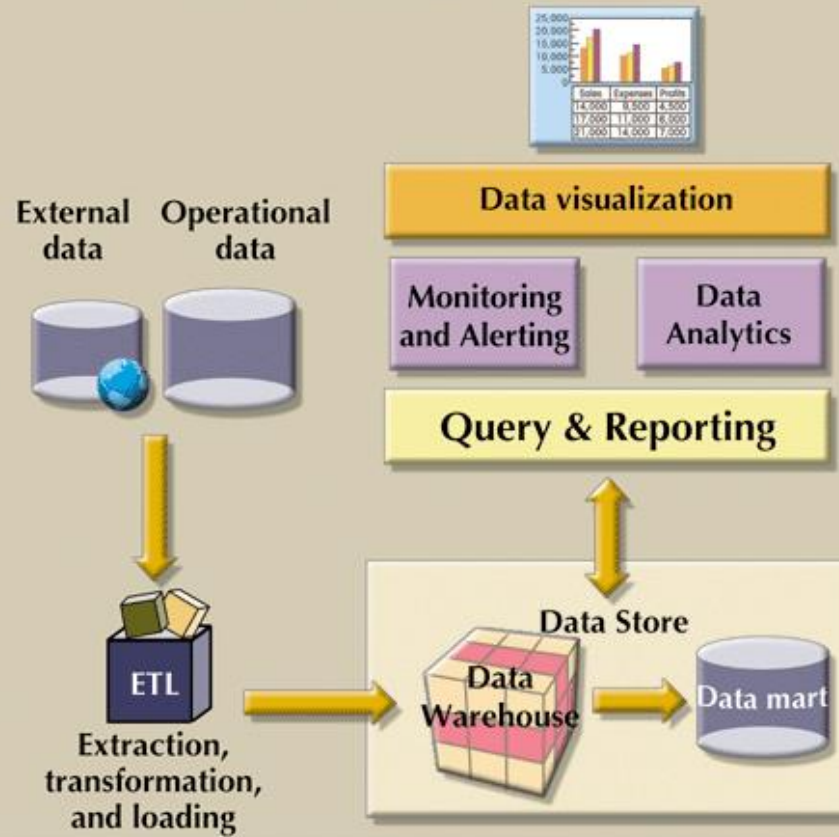
**FIGURE  
13.1**

## Business intelligence framework

People

Processes

### Business Intelligence Framework



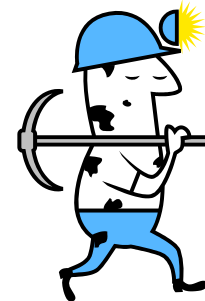
Management

Governance

SOURCE: Course Technology/Cengage Learning

## Ví dụ về Data Mining

- ❑ Nhiều năm trước, một nhà bán lẻ lớn hiện thực data warehouse để phân tích việc bán lẻ.
- ❑ Một khối lượng dữ liệu khổng lồ cỡ terabytes từ các điểm bán hàng (Point of Sale –POS) được đưa vào warehouse.
- ❑ Xây dựng ứng dụng, dựa trên phần mềm data mining chuyên dùng để thực hiện phân tích giỏ hàng (market basket analysis), trả lời câu hỏi:
  - Trong cùng 1 lần mua hàng, các món hàng nào được mua cùng với nhau?





## “Diapers and Beer”

- ❑ Để ý rằng trong 1 tương quan không bình thường chút nào, một trong số rất nhiều giao dịch (transactions) mua hàng lại có sự xuất hiện của bỉa và tã lót trẻ em trong cùng một giỏ hàng.
- ❑ Phân tích chỉ ra rằng có một phân khúc nhỏ của khách hàng, đó là các ông bố trẻ đi mua tã cho con, rồi quyết định mua thêm bia cho mình!

Luật kết hợp: Diaper  $\rightarrow$  Beer [0.5%, 75%]\*

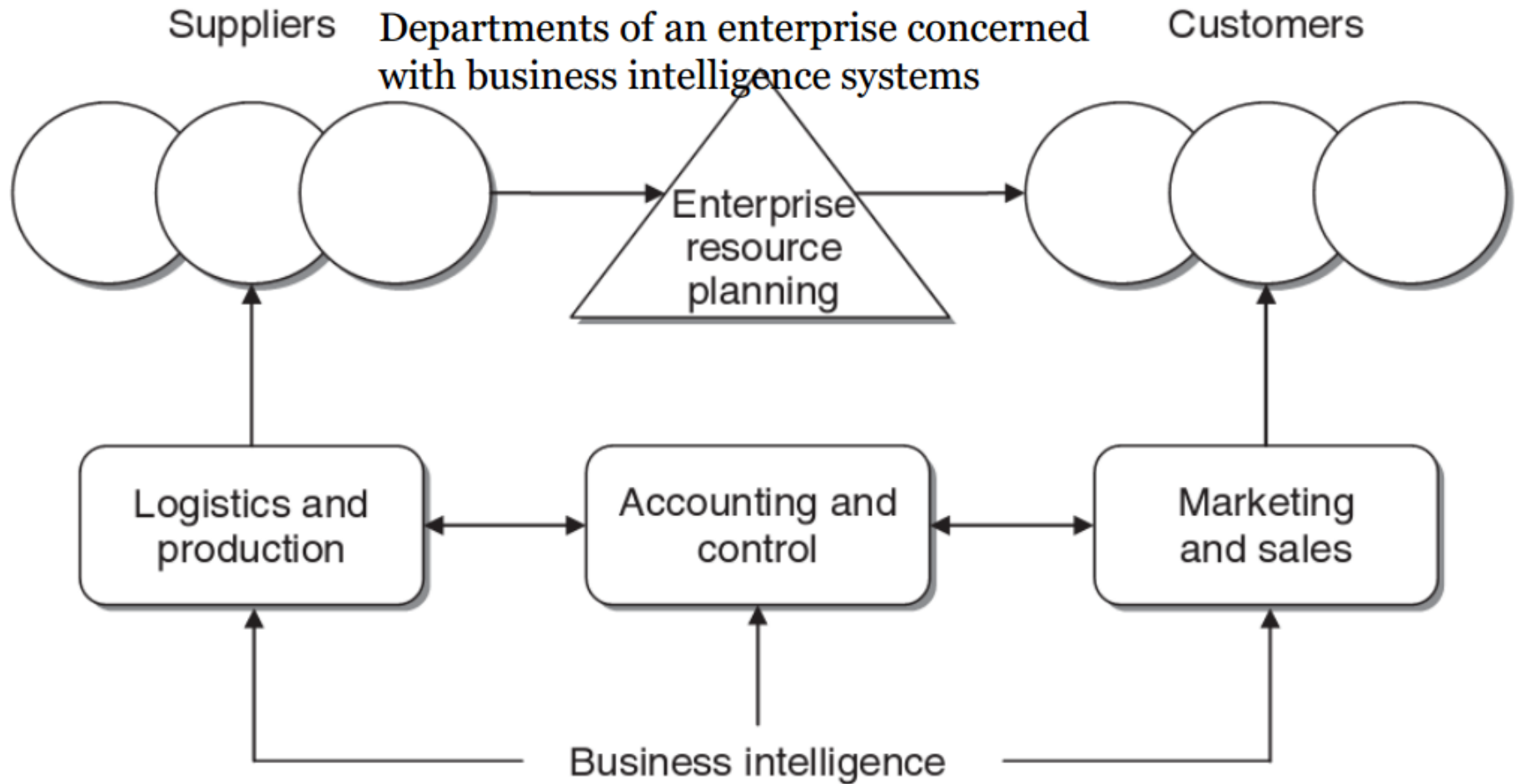
- ❑ Dựa trên thông tin này, nhà bán lẻ tổ chức lại quầy hàng bán tã: đặt thêm bia ở cuối quầy.
- ❑ **Doanh số bán bia tăng lên! – Bạn hiểu tác động của datamining vào bán hàng chưa?**

---

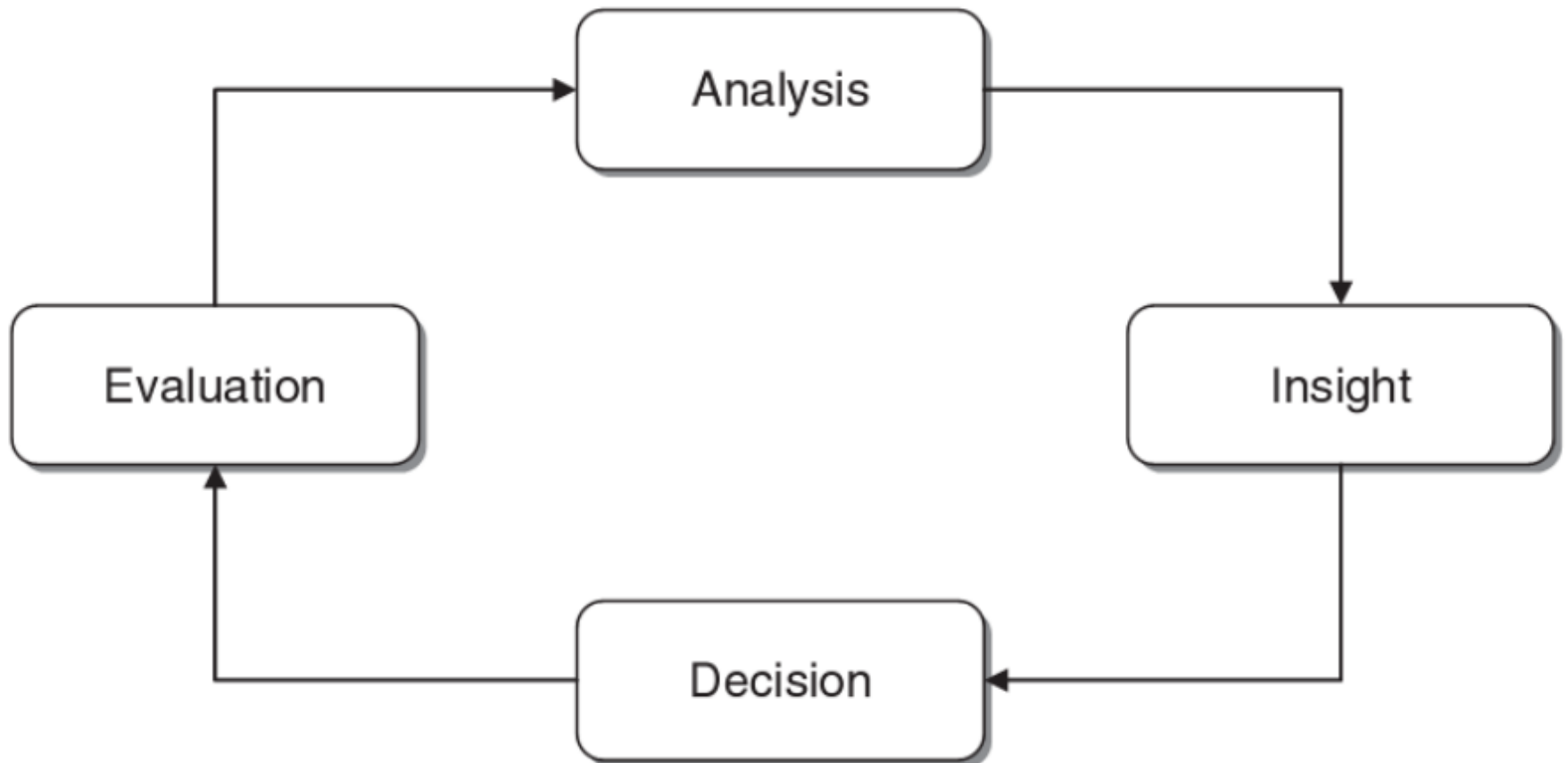
\*: 0.5%: độ hỗ trợ

75%: độ tin cậy

# Chu trình phân tích thông minh doanh nghiệp



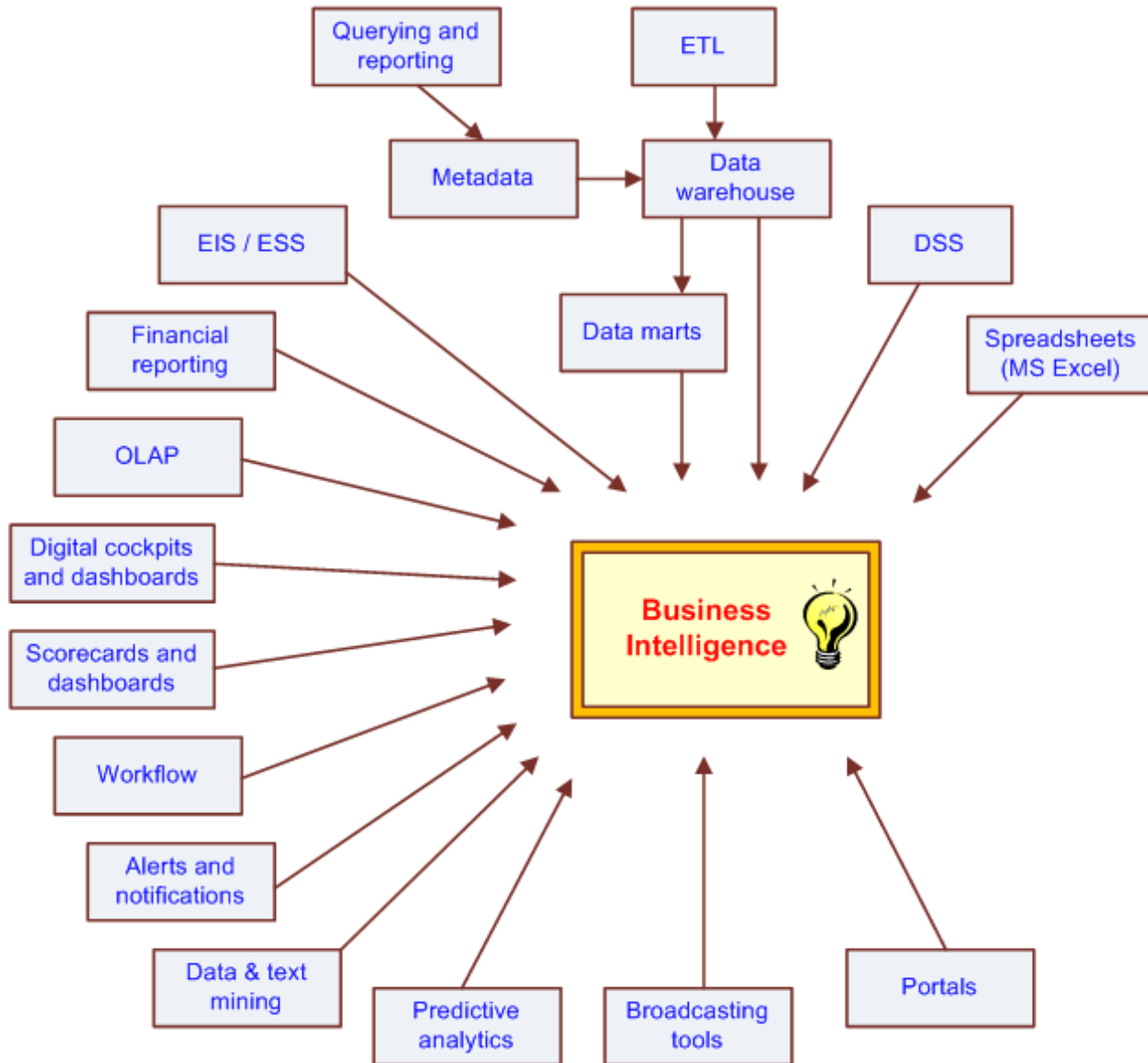
# Chu trình phân tích thông minh doanh nghiệp



# Khuyñh hướng công nghệ của BI

- ❑ Data storage improvements
- ❑ Business intelligence appliances
- ❑ Business intelligence as a service
- ❑ Big Data analytics
- ❑ Personal analytics

# The Evolution of BI Capabilities



---

**Tự đọc thêm**  
**Text and Web Mining**

---

# Text Mining Concepts

- ❑ 85-90 percent of all corporate data is in some kind of unstructured form (e.g., text)
- ❑ Unstructured corporate data is doubling in size every 18 months
- ❑ Tapping into these information sources is not an option, but a need to stay competitive
- ❑ Answer: text mining
  - A semi-automated process of extracting knowledge from unstructured data sources
  - As known as, text data mining or knowledge discovery in textual databases

# Data Mining versus Text Mining

- ❑ Both seek for novel and useful patterns
- ❑ Both are semi-automated processes
- ❑ Difference is the nature of the data:
  - Structured versus unstructured data
  - **Structured data:** in databases
  - **Unstructured data:** Word documents, PDF files, text excerpts, XML files, and so on
- ❑ Text mining – first, impose structure to the data, then mine the structured data



# Text Mining Concepts

- ❑ Benefits of text mining are obvious especially in text-rich data environments
  - e.g., law (court orders), academic research (research articles), finance (quarterly reports), medicine (discharge summaries), biology (molecular interactions), technology (patent files), marketing (customer comments), etc.
- ❑ Electronic communication records (e.g., Email)
  - Spam filtering
  - Email prioritization and categorization
  - Automatic response generation

# Text Mining Application Area

- ☐ Information extraction
- ☐ Topic tracking
- ☐ Summarization
- ☐ Categorization
- ☐ Clustering
- ☐ Concept linking
- ☐ Question answering

# Natural Language Processing (NLP)

## ❑ Structuring a collection of text

- Old approach: bag-of-words
- New approach: natural language processing

## ❑ NLP is ...

- a very important concept in text mining
- a subfield of artificial intelligence and computational linguistics
- the studies of "understanding" the natural human language

## ❑ Syntax versus semantics based text mining

# Natural Language Processing (NLP)

## ❑ What is “Understanding” ?

- Human understands, what about computers?
- Natural language is vague, context driven
- True understanding requires extensive knowledge of a topic
- Can/will computers ever understand natural language the same/accurate way we do?

# Natural Language Processing (NLP)

## ❑ Challenges in NLP

- Part-of-speech tagging
- Text segmentation
- Word sense disambiguation
- Syntax ambiguity
- Imperfect or irregular input
- Speech acts

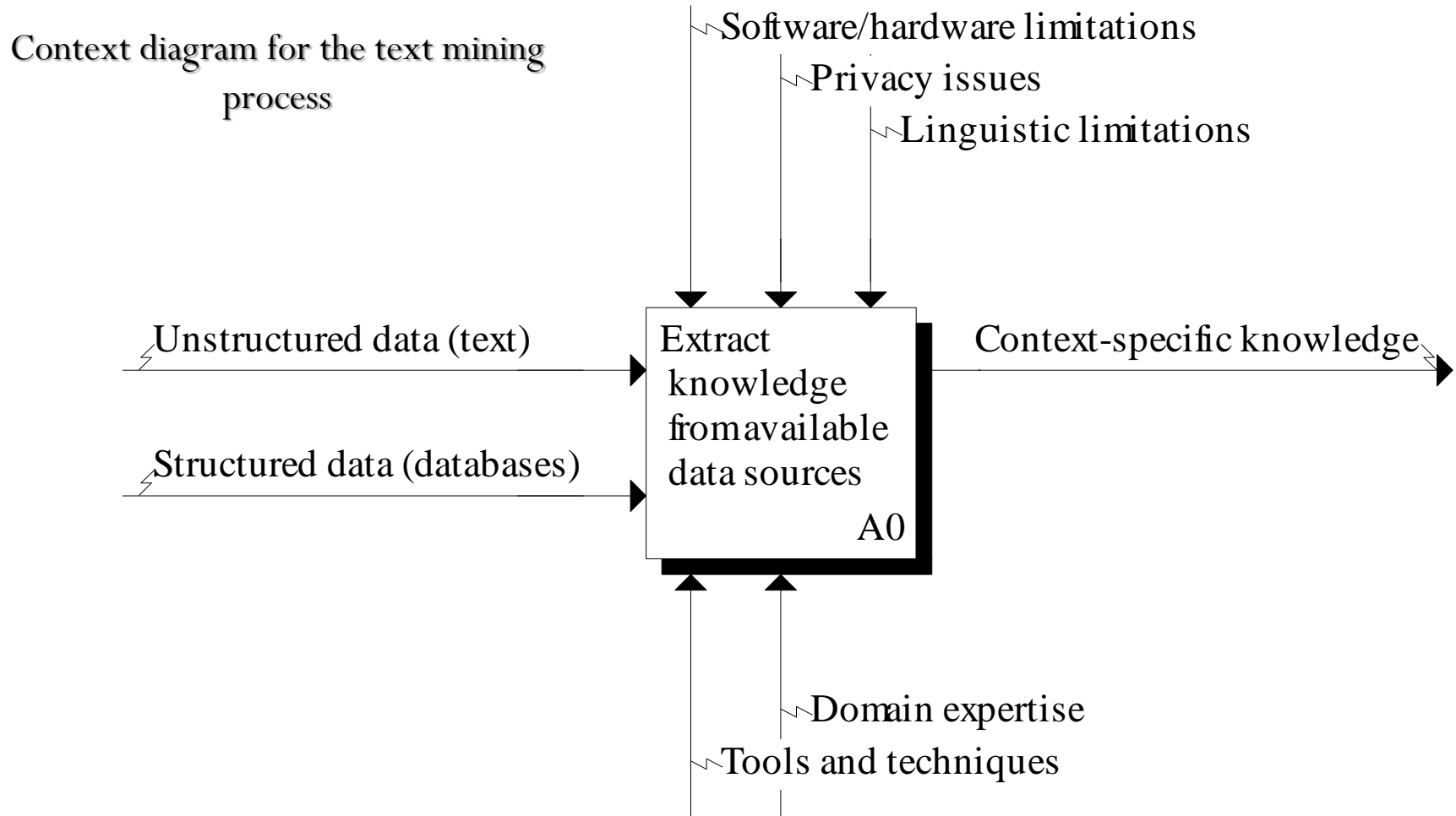
## ❑ Dream of AI community

- to have algorithms that are capable of automatically reading and obtaining knowledge from text

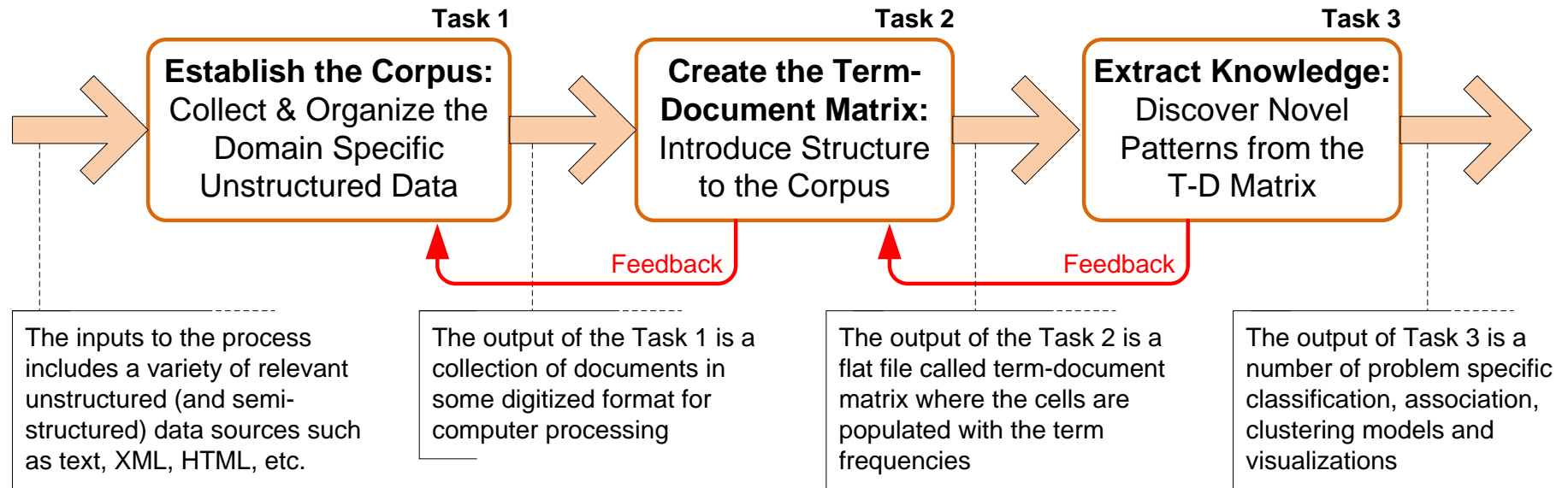
# NLP Task Categories

- ❑ Information retrieval
- ❑ Information extraction
- ❑ Named-entity recognition
- ❑ Question answering
- ❑ Automatic summarization
- ❑ Natural language generation and understanding
- ❑ Machine translation
- ❑ Foreign language reading and writing
- ❑ Speech recognition
- ❑ Text proofing
- ❑ Optical character recognition

# Text Mining Process



# Text Mining Process



The three-step text mining process



# Text Mining Process

## ❑ Step 1: Establish the corpus

- Collect all relevant unstructured data (e.g., textual documents, XML files, emails, Web pages, short notes, voice recordings...)
- Digitize, standardize the collection (e.g., all in ASCII text files)
- Place the collection in a common place (e.g., in a flat file, or in a directory as separate files)

# Text Mining Process

## □ Step 2: Create the Term-by-Document Matrix

<div>Terms</div> <div>Documents</div>	investment risk	project management	software engineering	development	SAP	...
Document 1	1			1		
Document 2		1				
Document 3			3		1	
Document 4		1				
Document 5			2	1		
Document 6	1			1		
...						

# Text Mining Process

❑ **Step 2:** Create the Term-by-Document Matrix (TDM), cont.

- Should all terms be included?
  - Stop words, include words
  - Synonyms, homonyms
  - Stemming
- What is the best representation of the indices (values in cells)?
  - Row counts; binary frequencies; log frequencies;
  - Inverse document frequency

# Text Mining Process

❑ **Step 2:** Create the Term-by-Document Matrix (TDM),  
cont.

- TDM is a sparse matrix. How can we reduce the dimensionality of the TDM?
  - Manual - a domain expert goes through it
  - Eliminate terms with very few occurrences in very few documents (?)
  - Transform the matrix using singular value decomposition (SVD)
  - SVD is similar to principle component analysis

# Text Mining Process

## ❑ Step 3: Extract patterns/knowledge

- Classification (text categorization)
- Clustering (natural groupings of text)
  - Improve search recall
  - Improve search precision
  - Scatter/gather
  - Query-specific clustering
- Association
- Trend Analysis (...)

# Text Mining Tools

## ❑ Commercial Software Tools

- SPSS PASW Text Miner
- SAS Enterprise Miner
- Statistica Data Miner
- ClearForest, ...

## ❑ Free Software Tools

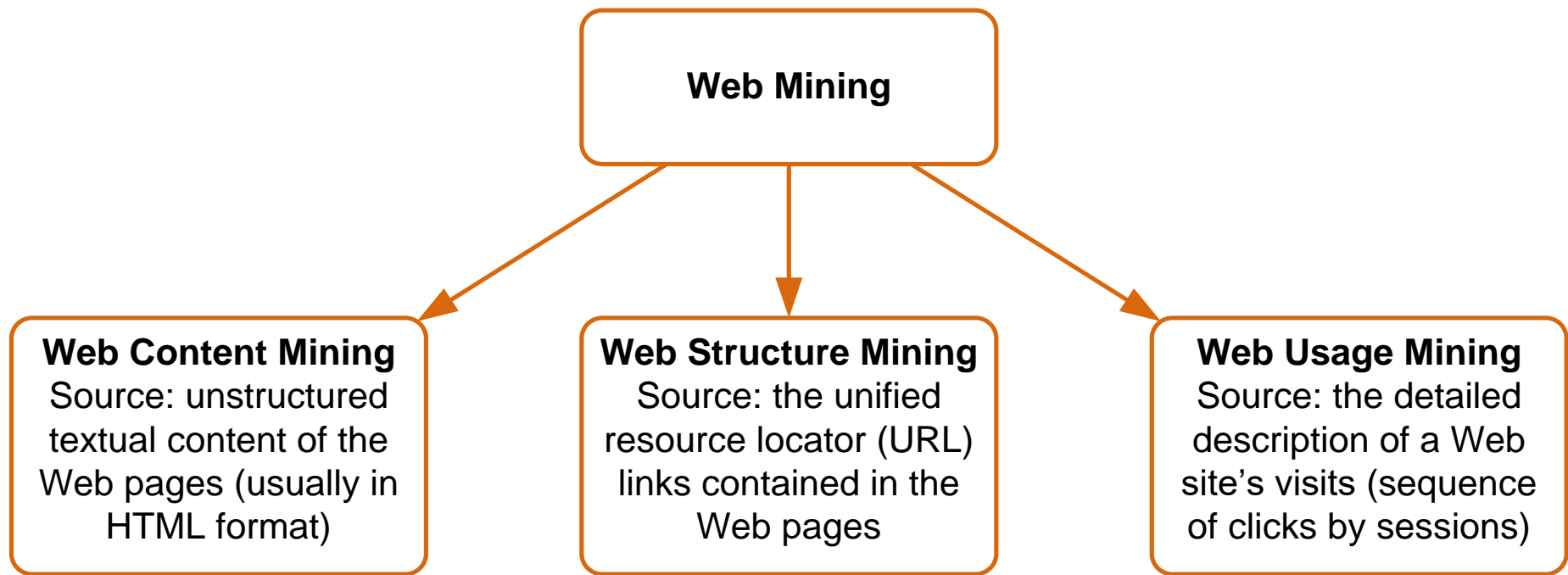
- RapidMiner
- GATE
- Spy-EM, ...

# Web Mining Overview

- ❑ Web is the largest repository of data
- ❑ Data is in HTML, XML, text format
- ❑ Challenges (of processing Web data)
  - The Web is too big for effective data mining
  - The Web is too complex
  - The Web is too dynamic
  - The Web is not specific to a domain
  - The Web has everything
- ❑ Opportunities and challenges are great!

# Web Mining

- ❑ Web mining (or Web data mining) is the process of discovering intrinsic relationships from Web data (textual, linkage, or usage)





# Web Content/Structure Mining

- ❑ Mining of the textual content on the Web
- ❑ Data collection via Web crawlers
  
- ❑ Web pages include hyperlinks
  - Authoritative pages
  - Hubs
  - Hyperlink-induced topic search (HITS)

# Web Usage Mining

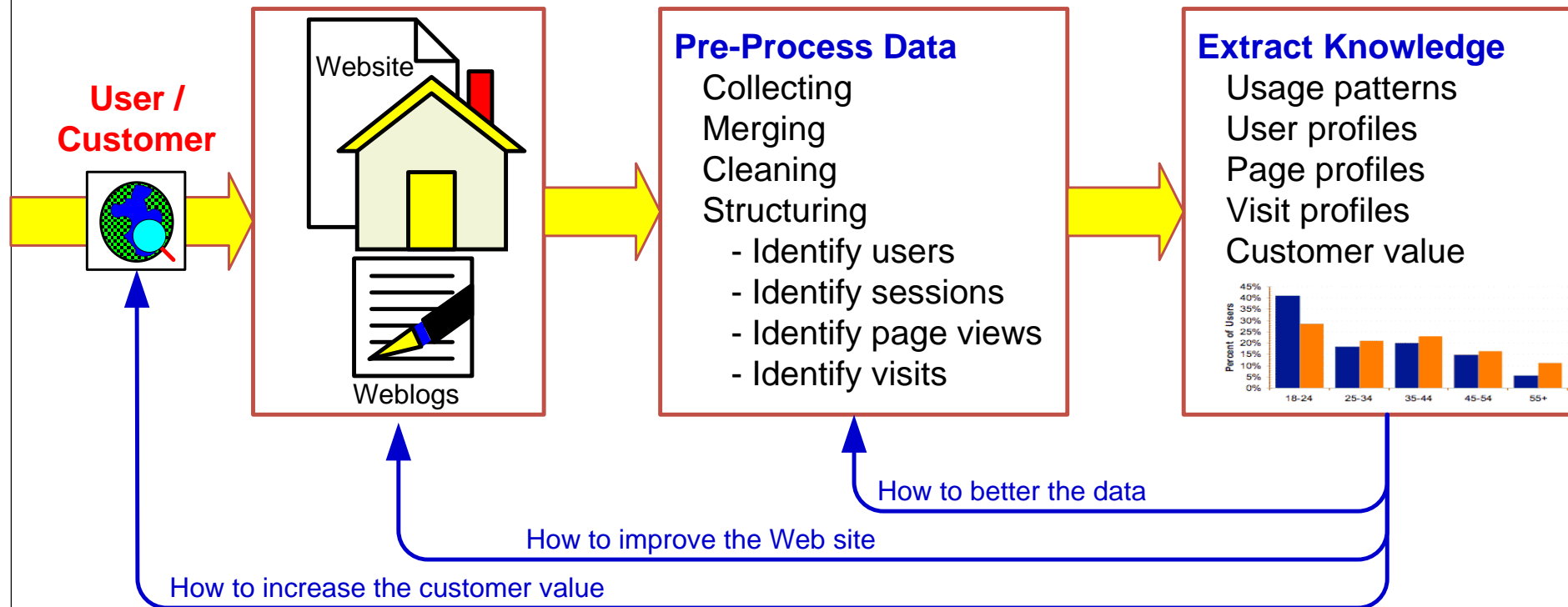
- ❑ Extraction of information from data generated through Web page visits and transactions...
  - data stored in server access logs, referrer logs, agent logs, and client-side cookies
  - user characteristics and usage profiles
  - metadata, such as page attributes, content attributes, and usage data
- ❑ Clickstream data
- ❑ Clickstream analysis

# Web Usage Mining

## ❑ Web usage mining applications

- Determine the lifetime value of clients
- Design cross-marketing strategies across products.
- Evaluate promotional campaigns
- Target electronic ads and coupons at user groups based on user access patterns
- Predict user behavior based on previously learned rules and users' profiles
- Present dynamic information to users based on their interests and profiles...

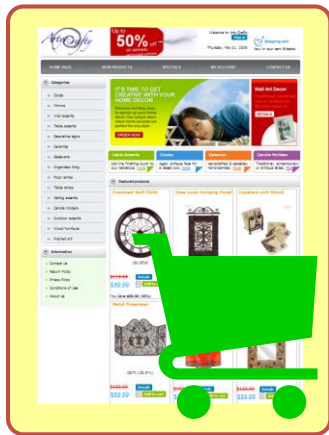
# Web Usage Mining (clickstream analysis)



# Web Mining Success Stories

- ❑ Amazon.com, Ask.com, Scholastic.com, ...
- ❑ Website Optimization Ecosystem

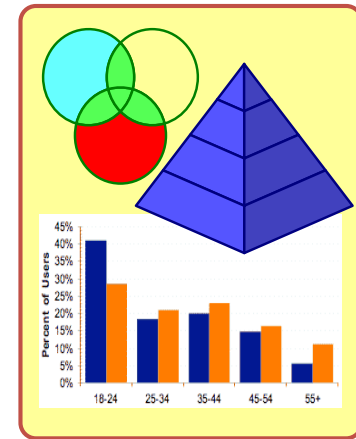
## Customer Interaction on the Web



## Analysis of Interactions



## Knowledge about the Holistic View of the Customer



# Mining the Social Web: Analyzing Data from Facebook, Twitter, LinkedIn, and Other Social Media Sites

---

*Analyzing Data from Facebook, Twitter, LinkedIn,  
and Other Social Media Sites*



Mining the  
Social Web

O'REILLY®

*Matthew A. Russell*

# Web Mining Tools

---

## Product Name

## URL

Angoss Knowledge WebMiner

[angoss.com](http://angoss.com)

ClickTracks

[clicktracks.com](http://clicktracks.com)

LiveStats from DeepMetrix

[deepmetrix.com](http://deepmetrix.com)

Megaputer WebAnalyst

[megaputer.com](http://megaputer.com)

MicroStrategy Web Traffic Analysis

[microstrategy.com](http://microstrategy.com)

SAS Web Analytics

[sas.com](http://sas.com)

SPSS Web Mining for Clementine

[spss.com](http://spss.com)

WebTrends

[webtrends.com](http://webtrends.com)

XML Miner

[scientio.com](http://scientio.com)

---