

# ***ETL***

## ***Extract-Transform-Load***

Prof. Sandro Jerônimo de Almeida



# ***Business Intelligence***

## **Definição**

- Conjunto de conceitos e metodologias que, fazendo uso de dados e acontecimentos (fatos), apoia a tomada de decisões em negócios
- Conjunto de técnicas que permite às empresas extrair informação útil, relevante, economicamente importante, previamente desconhecida, não-trivial a partir de múltiplas e volumosas fontes de dados brutos ou Sistematizados



# ***Business Intelligence***

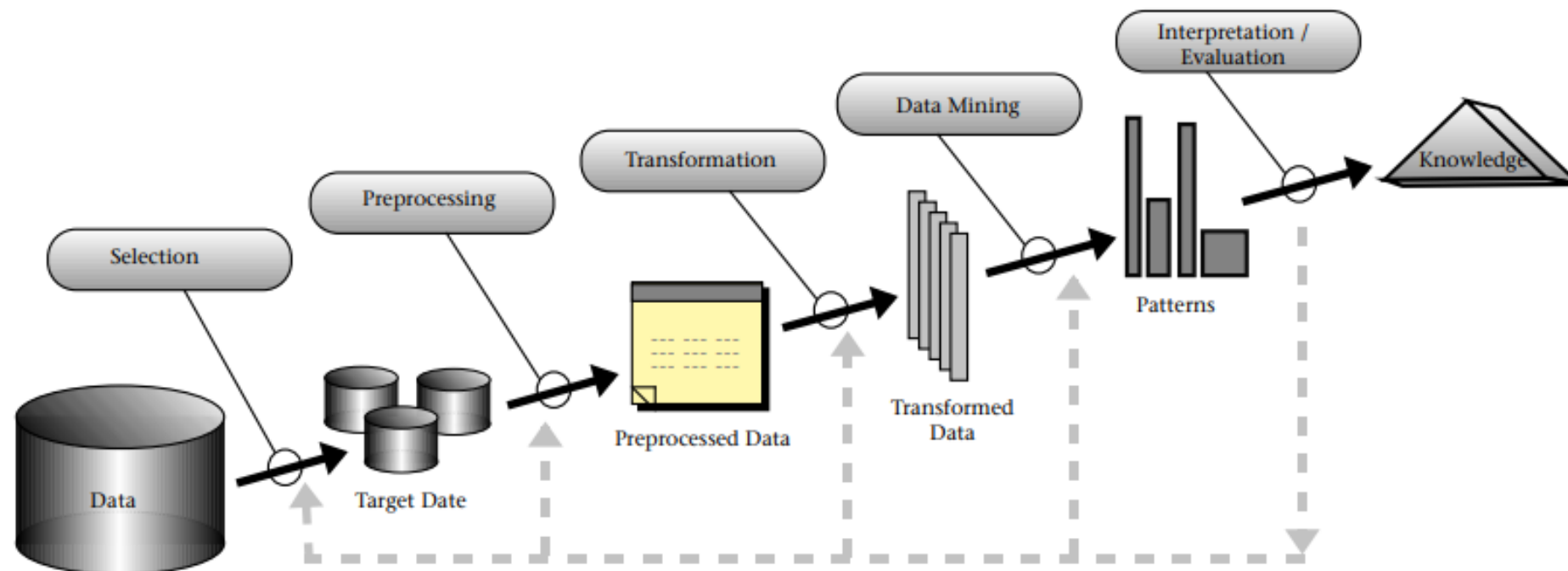
## Alguns elementos e tecnologias

- Relatórios
- Ferramentas de processamento analíticos (detalhados)
- Gerenciamento de Indicadores de desempenho
- Análises descritiva e Preditiva
- Técnicas de IA (ex. *Data Mining, Machine Learning*)
- *Visualização de dados*



# Processo de Descoberta de Conhecimento

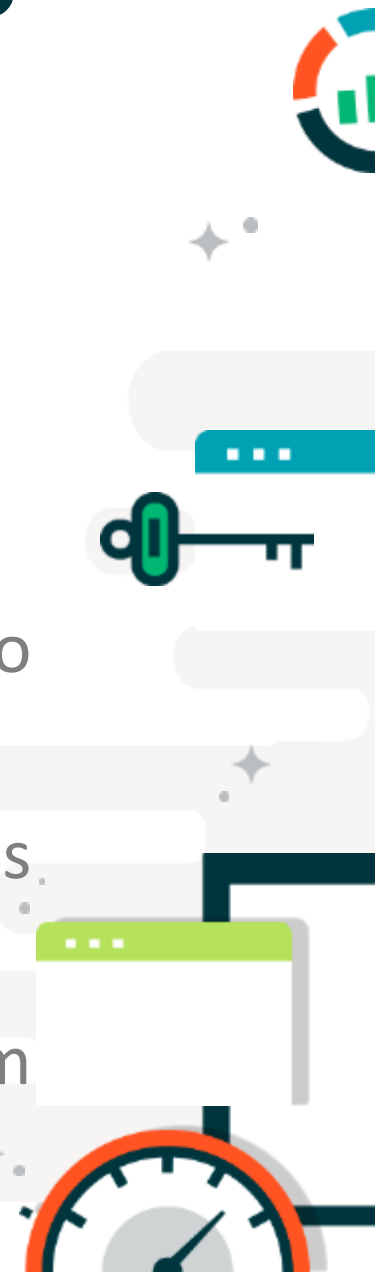
## KDD – Knowledge Discovery in Databases



**Fonte:** FAYYAD, U. M., Piatetsky Shapiro, G., Smyth, P. & Uthurusamy, R. "Advances in Knowledge Discovery and Data Mining" 1996, AAAIPress, The Mit Press.

# *ETL – Extração, Transformação e Carga*

- Facilita a importação, exportação e transformação de dados heterogêneos
- Integração , consolidação e agrupamento de informações
- Pode-se utilizar ferramentas comerciais ou seu próprio algoritmo/*script*
- Áreas intermediárias podem ser necessárias para manipulação dos dados (*staging area*)
- É comum levar até 80% do tempo no processo de ETL do que em outras etapas da busca por conhecimento em base de dados.



# Processo ETL

## Fontes de Dados



# *ETL – Extração, Transformação e Carga*

## Limpeza de Dados

- Como o DW é usado para tomada de decisão, é importante que os seus dados estejam corretos.
- Entretanto, uma vez que grandes volumes de dados estão envolvidos, há uma alta probabilidade de erros e anomalias nos dados.
- Tamanhos inconsistentes de campo, descrições inconsistentes, atribuição inconsistente de valores, entradas erradas e violação de restrições de integridade são alguns exemplos onde a limpeza de dados torna-se necessária.



# *ETL – Extração, Transformação e Carga*

## Transformação de Dados

- Conjunto de operações aplicado a cada registro ou linha de uma tabela antes de ser enviada ao seu destino
- Exemplo 1: cópia de uma coluna para outra com nome diferente (cod\_cli, codCliente)
- Exemplo 2: cálculo do valor destino baseado em múltiplas colunas de origem (qtd x valor)





# *ETL – Extração, Transformação e Carga*

## Carga de Dados

- Após extrair, limpar e transformar, os dados devem ser carregados para um repositório de destino.
- Um pré-processamento adicional pode ser requerido, como por exemplo, checagem de restrições de integridade, sumarização, agregação, dentre outros mais.
- Tipicamente o processo de carga é feito em lotes. Essa carga geralmente tem que lidar com volumes de dados muito grandes.



