

# CONCURSEIRO

## FORA DA CAIXA

INFORMÁTICA E TI  
BRINDE - TÓPICOS SELECIONADOS

HENRIQUE DE LARA MORAIS  
[www.concursseiroforadacaixa.com.br](http://www.concursseiroforadacaixa.com.br)

# Sumário

---

<b>Informática .....</b>	<b>3</b>
Hardware .....	3
CPU – Central process Unity .....	3
Memórias.....	3
Placa mãe.....	4
Redes.....	4
Dimensão .....	4
Comutação de Circuito e Comutação de Pacote .....	4
Topologias de Rede .....	4
Internet, Intranet e Extranet.....	5
Equipamentos de Rede.....	6
Segurança.....	6
<b>Tecnologia da Informação.....</b>	<b>7</b>
Análise de Dados.....	7
Dado, Informação, Conhecimento e Inteligência .....	7
Banco de Dados e SGBD .....	7
Nível de Abstração de Dados .....	8
Modelagem de Dados.....	9
Data Mart e Data Warehouse .....	11
Data Warehouse (DW) .....	11
Data Mart (DM) .....	11
Abordagens para Projeto de Datawarehouse.....	11
ETL – Extract, Transform, Load.....	11
Modelagem Dimensional .....	12
Segurança da Informação.....	13
Princípios .....	13
Criptografia.....	14
Assinatura Digital.....	14
Certificado Digital.....	15
Algumas pragas bastante cobradas .....	15
Escrituração Fiscal Digital – EFD.....	16
EFD – Blocos e Registros .....	16
Nota Fiscal Eletrônica – NFE .....	17

## Observações

---

Antes de você utilizar esse material é importante fazer uma observação. Informática e TI são assuntos gigantescos, e o nível de detalhes e possibilidades de cobrança beira o infinito. Não tenho a intenção de vender esse material pelo simples fato de acreditar que vocês merecem algo melhor.

TI e Informática serão cada vez mais cobrados e com dificuldade cada vez maior. Pensando nisso, acho que seria legal montarmos um **material colaborativo**. Como assim? Bom, vou deixar um formulário abaixo para VOCÊ preencher com um ou mais tópicos e uma ou mais sugestões de resumo do que **você acha relevante**. Por exemplo:

1. Tópico: Gestão de TI – ITIL
2. Sugestão: aqui você coloca SEU resumo (algo que você já fez), como por exemplo, a “Matriz RACI”.

Dessa forma, vou tentar compilar todas as sugestões 1x por mês, de modo que tenhamos um excelente material para revisão, **feito por todos e para todos, sem custo**. Quanto à figuras e esquemas pode deixar que eu dou um jeito. Para acessar o formulário basta clicar no botão abaixo:



Para finalizar: eu coloquei a parte genérica de EFD e NFe aqui, por ter sido cobrado no ICMS-SC. Isso vai especialmente para o pessoal que estuda para a **área fiscal**, pois me parece ser uma tendência.

Abraços,  
Henrique

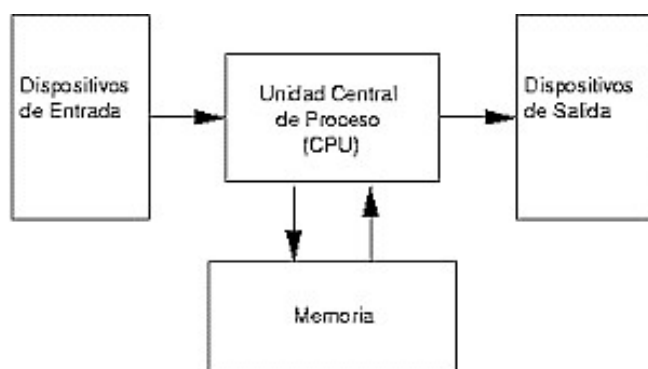
## INFORMÁTICA

### HARDWARE

Esse tema tem muitos tópicos que fazem parte do nosso cotidiano. Falar em mouse, teclado, impressora, scanner, etc. é bastante comum, por isso creio que as questões não trarão muitas dificuldades para você quando envolver esses itens. Para ser bem objetivo e não ficar ocupando seu tempo de revisão com bobagem, resolvi colocar aqui apenas aqueles componentes que realmente podem causar alguma dúvida ou confusão.

### CPU – CENTRAL PROCESS UNIT

**Processador:** é chamado de CPU (unidade central de processamento) e está acoplado à placa-mãe. Faz **todo o controle das operações** que serão realizadas pelo computador. Arquitetura de Von Neumann:



CPU é dividida em:

- Unidade Lógica e Aritmética (ULA): executa as operações matemáticas propriamente ditas;
- UC (unidade de controle): instrui para onde nas memórias a informação deve ir
- Registradores: são memórias que **trabalham no ritmo dos processadores**. São **VOLÁTEIS**.

Processador 32 bits:  $2^{32} = 4\text{GB}$

Processador 64 bits:  $2^{64} = 16\text{GB}$

### MEMÓRIAS



<u>RAM (memória principal)</u>	<u>ROM (memória principal)</u>	<u>Memória Secundária</u>
<ul style="list-style-type: none"> <li>• <b><u>Volátil e rápida</u></b></li> <li>• Muito <b>mais cara</b></li> <li>• Só funciona com PC ligado</li> <li>• Uma vez desligado, ela “limpa”</li> <li>• CPU a utiliza para armazenar <u>temporariamente</u> os dados dos programas que estão rodando.</li> </ul>	<ul style="list-style-type: none"> <li>• Armazenamento <b>permanente</b></li> <li>• Tecnologia: <b>CMOS</b></li> <li>• Dados não apagam ou se alteram</li> <li>• Dados não se perdem ao desligar (precisam da pilha)</li> <li>• Ex: flash, CD-ROM, DVD-ROM</li> </ul>	<ul style="list-style-type: none"> <li>• Pen drives, CD, DVD, HD, etc.</li> <li>• Retém grande quantidade</li> <li>• São <b>NÃO</b> voláteis</li> </ul>

**Memória Cache:** memória de acesso randômico **VOLÁTIL** mais rápida que **armazena os dados mais utilizados pelo processador (evita ficar acessando toda a hora a RAM, que é lenta)**. Sem a memória cache o desempenho da máquina ficaria mais lento e limitado à memória RAM. Existem dois tipos atualmente:

- Cache de nível 1 (cache L1) - localizada no mesmo chip do processador;
- Cache nível 2 (cache L2) - localizada geralmente em um chip RAM separado.

**Memória Virtual:** é uma técnica que permite a simulação da existência de mais memória RAM do que o micro realmente tem. Essa simulação é feita no disco rígido.

## PLACA MÃE

**BIOS (= firmware):** um **SOFTWARE, gravado em um chip de memória ROM** (que fica espetado na placa-mãe do computador). Trata-se de um sistema responsável por iniciar os trabalhos de um computador. Ele checa, por exemplo, o estado das memórias e verifica a presença de dispositivos de E/S, em seguida, faz a carga do sistema operacional no disco (rígido ou flexível), entregando o controle ao sistema operacional.

**Barramento:** também chamado de bus são suportes responsáveis por fazer a intercomunicação entre a placa-mãe e os demais componentes.

**CHIPSET:** constitui o **conjunto principal de circuitos integrados**, aqueles responsáveis por controlar todo o fluxo de dados na placa mãe. **Já vem instalado nativamente em qualquer placa mãe.**

## REDES

Conforme comentei nas observações iniciais, não tenho a pretensão de me aprofundar em qualquer um dos temas de Informática e/ou TI. Vou apenas pincelar aqui algumas considerações.

## DIMENSÃO

**Rede:** conexão de **DOIS ou MAIS computadores** para permitir o compartilhamento de recursos e a troca de informações entre as máquinas;

**LAN (Local Area Network):** é uma rede de uma casa, escritório, ou seja, “pequena” – **área de até 10Km**

**WLAN (Wireless LAN):** é uma LAN sem fio, por sinais de rádio (WiFi);

**CAN (Campus Area Network):** interconecta computadores (e/ou outros dispositivos) de vários edifícios;

**MAN (Metropolitan Area Network):** é uma rede grande o bastante para não ser uma rede meramente local;

**WAN (Wide Area Network):** qualquer rede que exceda o tamanho de uma MAN – EX: várias cidades, um país, o mundo inteiro (internet);

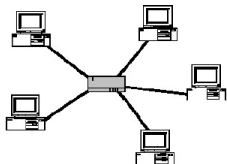
## COMUTAÇÃO DE CIRCUITO E COMUTAÇÃO DE PACOTE

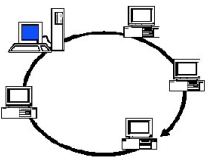

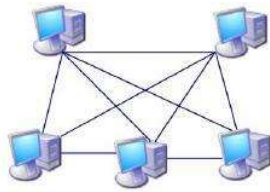
**Comutação de Circuitos (ou por Circuitos)** sempre existe uma fase prévia à transmissão de dados em si, que é o estabelecimento do circuito. Pense numa ligação telefônica, por exemplo, antes de você começar a falar, **existe a necessidade de estabelecer a ligação, o circuito**, entre você e seu destinatário. É **necessária uma ligação física entre os pontos, por um caminho ÚNICO**. É uma tipologia antiga.

**Comutação de Mensagens e de Pacotes**, cada mensagem ou pacote é enviado diretamente na rede, sem o estabelecimento prévio de um circuito. Para isso, cada unidade de comunicação (mensagem ou pacote) precisa estar adequadamente etiquetada com o destinatário, e **a rede irá fazer com que cada um chegue ao mesmo destino**. Observe que **é possível que o caminho ao longo da rede se modifique entre a transmissão de um pacote e a do outro**.

Outra analogia possível é que a comutação de circuito seria um trem, por meio do qual vários passageiros chegarão a um certo destino trilhando o mesmo caminho. Já na comutação de mensagens ou pacotes os passageiros seguem em diversos carros de passeio, podendo trilhar caminhos diferentes.

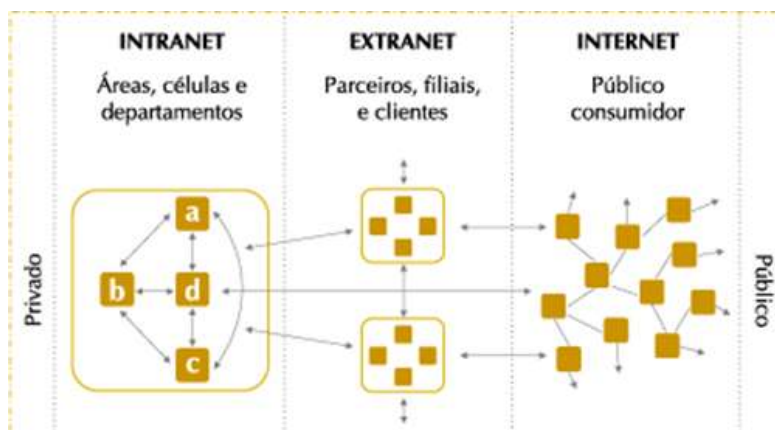
## TOPOLOGIAS DE REDE

TOPOLOGIA	PONTOS POSITIVOS	PONTOS NEGATIVOS
<p><b>Estrela</b></p> 	<ul style="list-style-type: none"> <li>• A falha de um PC não afeta os demais.</li> <li>• Fácil acrescentar novos PC's.</li> <li>• <b>Gestão CENTRALIZADA.</b></li> <li>• PC que quiser falar com outro, manda para nó central e este encaminha ao PC</li> </ul>	<ul style="list-style-type: none"> <li>• Custo de instalação maior pois recebe mais cabos.</li> <li>• Se o ponto central falha, a rede falha.</li> </ul>

TOPOLOGIA	PONTOS POSITIVOS	PONTOS NEGATIVOS
<p><b>Anel</b></p> 	<ul style="list-style-type: none"> <li>A mensagem de um atravessa todo o anel.</li> <li>Requer <b>menos cabos (BAIXO CUSTO)</b></li> <li>Desempenho uniforme.</li> </ul>	<ul style="list-style-type: none"> <li>Os problemas são difíceis de isolar.</li> <li>Se um parar de funcionar, <b>toda a rede cai</b>, sendo difícil identificar o ponto onde ocorreu a falha.</li> </ul>
<p><b>Barra (linear)</b></p> 	<ul style="list-style-type: none"> <li>Simple e fácil de instalar.</li> <li>Fácil de ampliar.</li> <li>Requer <b>menos cabos (BAIXO CUSTO)</b></li> </ul>	<ul style="list-style-type: none"> <li>A rede funciona por difusão (broadcast).</li> <li>Rede mais lenta em períodos de uso intenso</li> <li>Colisão de dados (literalmente de eletricidade)</li> <li>Os problemas são difíceis de isolar.</li> </ul>
<p><b>Malha (full meshed)</b></p> 	<ul style="list-style-type: none"> <li>A falha em um nó não prejudica os demais.</li> <li>Topologia livre de colisões.</li> <li>Melhor desempenho entre todas as topologias.</li> <li>A falha em uma ligação não impede a comunicação com os demais nós.</li> <li>O aumento de nós não degrada o desempenho da rede.</li> </ul>	<ul style="list-style-type: none"> <li><b>Custo altíssimo</b>, uma vez que para uma rede com N nós, são necessários <math>N(N-1)/2</math> links.</li> </ul>

Atualmente, não se utiliza uma única topologia dentre as listadas. Utilizam-se topologias híbridas, ou seja, uma mistura de cada uma das topologias listadas de acordo com o custo ou a necessidade de desempenho

## INTERNET, INTRANET E EXTRANET



Internet: composta por **todos os computadores do mundo** ligados em rede. Seu funcionamento é **baseado na Pilha de Protocolos TCP/IP**;

Intranet: serviço análogo à Internet, **também baseada na pilha de protocolos TCP/IP**. Porém, a Intranet é **restrita a um local físico**;

Extranet: funciona como uma **extensão da Intranet** a computadores que estejam **fora dos domínios físicos da Intranet**.

**Importante**: a Intranet e Extranet, geralmente, também utilizam as mesmas tecnologias da Internet.

## EQUIPAMENTOS DE REDE



**MODEM** é um equipamento que modula um sinal digital numa onda analógica, pronta a ser transmitida pela linha telefônica, e que demodula o sinal analógico e reconverte-o para o formato digital original.

**ROTEADORES:** roteia pacotes (daí vem o nome). Rotear significa **escolher o melhor caminho na rede para o pacote seguir e chegar ao seu destino**. Enfim, fazem o controle de tráfego de dados da internet.

**SWITCH** (comutador): permite a conexão física de cabos provenientes de vários nós. É capaz de ler a informação que recebe e identificar seu destinatário, enviando-a APENAS ao nó destinatário (diferentemente do hub, que envia a todos os nós conectados ao equipamento).

**ACCESS POINT:** é o famoso repetido de sinal de Wi-Fi

## SEGURANÇA

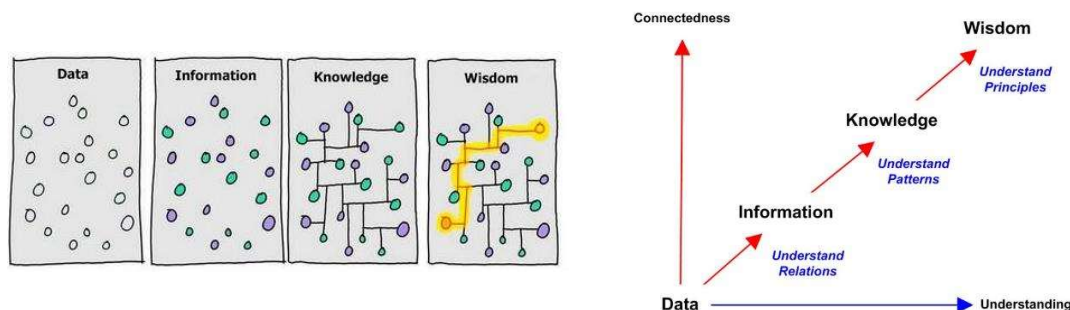
**WEP** (Wired Equivalent Privacy) foi criado em 1999. O padrão **WEP se torna mais inseguro à medida que o poder de processamento dos computadores aumenta**. Como utiliza chaves de 128 bits, a senha pode ser descoberta em pouquíssimo tempo

**WPA2** (Wireless Protected Access) utiliza protocolo **AES** (Advanced Encryption Standard), proporcionando alto grau de segurança e tornando redes Wi-Fi mais confiáveis. Chaves de 128 a 256 bits.

## TECNOLOGIA DA INFORMAÇÃO

### ANÁLISE DE DADOS

#### DADO, INFORMAÇÃO, CONHECIMENTO E INTELIGÊNCIA



**Dados:** é um **REGISTRO** de alguma entidade. Dados são fatos que podem ser analisados e que possuem um significado implícito. Estão na forma de símbolos, imagens, números, etc.

- **Estruturados:** formato **PADRONIZADO (rigidez)** p/ cada atributo. São mantidos em banco de dados. Ex: Coluna A (telefone, só nºs, até 9 caracteres) | Coluna B (nome, só letras, até 20 caracteres).
- **Semiestruturados:** **não existe esquema padrão** predefinido, sendo definido a posteriori, **após a existência dos dados** (é o famoso: preencha aqui \_\_\_\_).
- **NÃO estruturados:** não possuem estrutura definida. É a grande maioria dos dados encontrados. Ex: documentos, textos, imagens e vídeos.

**Informações:** é o resultado do **processamento, manipulação e organização de dados** de forma a terem significado. É um dado acrescido de **contexto, relevância e propósito**. Não há conclusão, mas organização que permite a análise.

**Conhecimento:** inclui a **reflexão, síntese e contexto** acerca da informação. É a análise da informação, sua relevância e importância, de forma a **assessorar a tomada de decisões**.

- **Conhecimento Tácito:** armazenado nas pessoas (experiência de vida);
- **Conhecimento Explícito / Codificado:** encontrado na forma de texto.

**Inteligência:** é o conhecimento **contextualmente** relevante que permite atuar (tomada de decisão) com vantagem no ambiente considerado. É o **conhecimento sintetizado e aplicado** a determinada situação.

#### BANCO DE DADOS E SGBD

**Banco de Dados:** **CONJUNTO de dados integrados / RELACIONADOS**, **logicamente coerente** (deve haver contexto), com a **finalidade de atender às necessidades de uma comunidade** de usuários (deve ter **fim específico**).

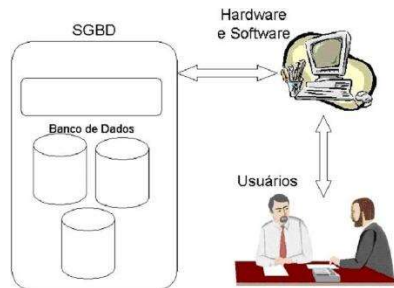
[DECORAR] **Requisitos para um BD (ACID)**

- **Atomicidade** – “tudo ou nada”, a transação ocorre ou não ocorre, não há meio termo
- **Consistência** – uma transação só pode sair de uma forma consistente para outra consistente
- **Isolamento** – garantia que na execução de uma transação, outra não realize uma interferência
- **Durabilidade** – garantia que uma transação feita não será perdida (não pode desfazer)

**SGBD:** **SOFTWARE** ou **CONJUNTO DE SOFTWARES** que possui recursos capazes de **armazenar, modificar e extrair e MANIPULAR** as informações do BD e **INTERAGIR** com o usuário. Exs: Oracle, SQL Server, DB2, PostgreSQL, MySQL, o próprio Access ou Paradox, entre outros.

- O SGBD lida com dados e com metadados (dados sobre os dados, indicando, por exemplo que o campo “idade” é um número inteiro, não negativo, que representa a idade de uma pessoa)
- Deve permitir uma série de operações básicas em um banco de dados, como a inserção, modificação, exclusão, pesquisa e ordenação de registros no BD (banco de dados)

Sistema de Banco de Dados: conjunto de quatro componentes básicos: dados, hardware, software e usuários. O sistema de banco de dados é "o todo". SBD = SGBD + BD



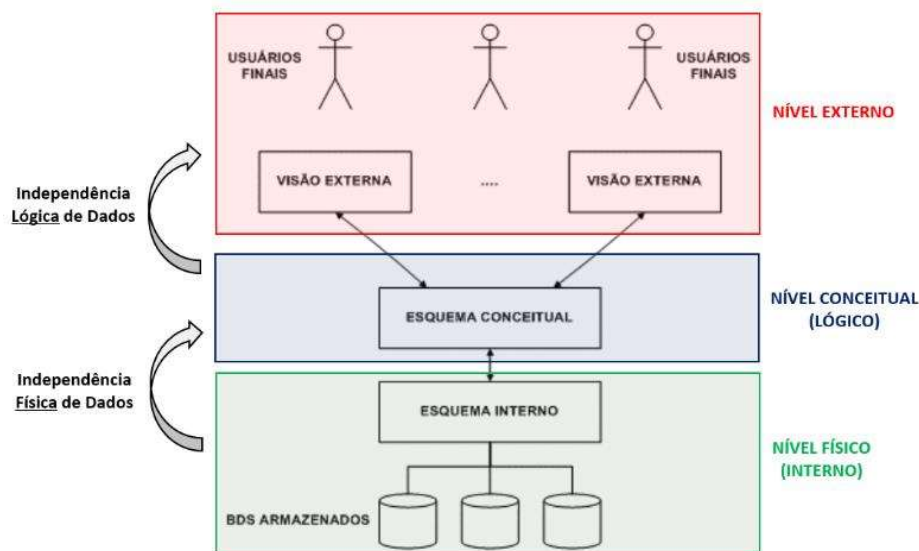
Dicionário de Dados: é uma coleção de metadados que contém definições e representações de elementos de dados. É um grupo de tabelas habilitadas **apenas para leitura ou consulta**, ou seja, é uma base de dados propriamente dita. Possui informações como por exemplo: em campos onde houver nº telefone, o formato deve ser (XX) XXXX-XXXX; campos com sexo só aceitam M ou F, etc.

## NÍVEL DE ABSTRAÇÃO DE DADOS

Nível de Visões do Usuário (Externo / View): nível **MAIS ALTO** de abstração, que descreve **PARTES do BD**, de acordo com as necessidades do usuário, i.e., descreve o **modo pelo qual os dados são VISTOS pelos usuários do SGBD**.

Nível Lógico (Conceitual): descreve **QUAIS dados estão armazenados e seus relacionamentos**. Faz a interface entre os dados físicos e as visões dos usuários.

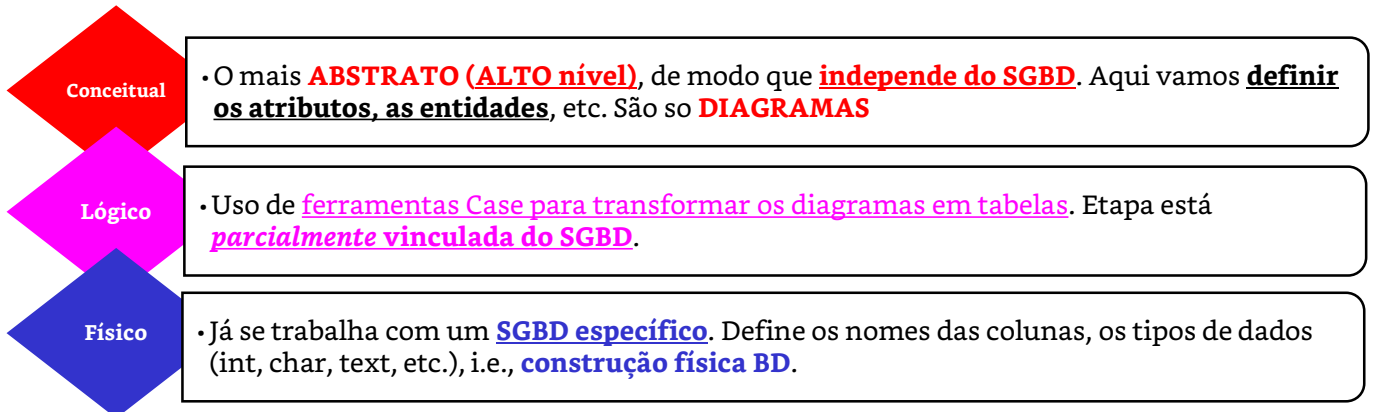
Nível Físico (Interno): Nível **mais baixo de abstração**. Descreve **COMO os dados estão realmente armazenados**, englobando estruturas complexas de baixo nível que são descritas em detalhe.



*Obs: independência é a capacidade de alterar um esquema sem alterar o seguinte.*


## MODELAGEM DE DADOS

### MODELO CONCEITUAL, LÓGICO E FÍSICO



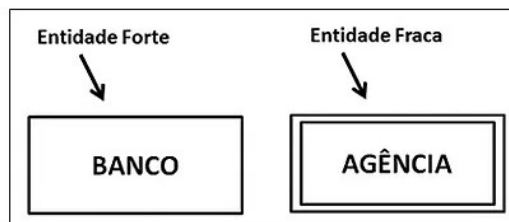
### MODELAGEM ENTIDADE-RELACIONAMENTO (MER)

No modelo relacional os dados são organizados em **coleções de tabelas Bidimensionais**. O Modelo ER **é um MODELO CONCEITUAL de ALTO nível**, cuja característica é refletir da **forma mais próxima possível a visão que o USUÁRIO tem dos dados**. Nesse modelo, **NÃO há preocupação em refletir como os dados estarão FISICAMENTE armazenados**.


**ENTIDADES**: algo significativo, sobre o qual devemos possuir informações. Exemplos: Cliente, Produto, Venda, Turma, Função, entre outros. Gráficamente: 

**Entidade Fraca**: está **vinculada à existência de outra entidade** (ela tem seus próprios atributos, mas DEPENDE de outra).

Gráficamente 



**TABELA**: estrutura básica de armazenamento no SGBR (R de relacional). Armazena todos os dados necessários sobre algo do mundo real. Um BD relacional pode ter uma ou mais tabelas. No **MER as tabelas são chamadas de relações**.

**ATRIBUTO / CAMPO**: unidade que armazena um tipo específico de dado (ou não armazena nada, como valor nulo). Nada mais é que uma COLUNA da tabela. Um exemplo seria a coluna "nome", "CPF", etc. Gráficamente: 





**Simples ou Atômicos**: atributos **não divisíveis**. Ex: CPF, RG, CEP, sexo, idade (se for para preencher)

**Composto**: podem ser **divididos em partes com significado diferente**. Ex: endereço (**sempre!**) – pode ser dividido em Rua, nº, Bairro, complemento, etc.

**Monovalorados**: possui apenas **um VALOR** p/ aquela entidade. Ex: nº da casa


**Multivalorados**: quando possui **diversos VALORES**. endereço (**sempre!**).

**Chave Primária (PK)**: coluna (atributo) que identifica um registro de **forma EXCLUSIVA** na tabela (relação), ou seja, **depende de quem estará usando a informação**. Ex: CPF, Endereço MAC, ISBN, Cód. Cliente, etc. A chave primária pode ser também um conjunto de atributos (EX: em um condomínio, a PK poderia ser "bloco" + "nº apartamento").

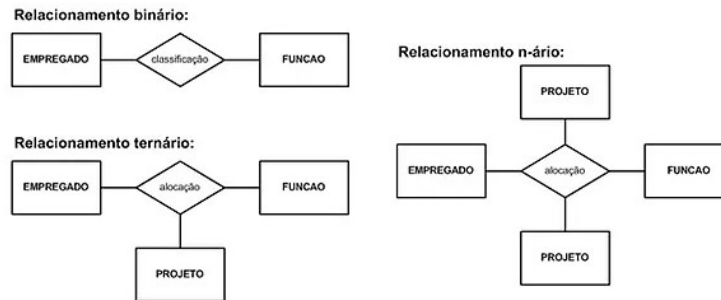
-  **Cuidado!** Quando uma PK se aplica a duas entidades, no diagrama costuma-se colocar no relacionamento (losango).
-  **A PK, obviamente, não pode ter valor vazio (NULL).**
-  **PODE existir +1 PK**
-  **Chave CANDIDATA / Alternativa**: é uma **coluna que poderia ser PK dentro de uma MESMA tabela** (ex: tabela com CPF e CodFuncionario. Se CPF for PK, CodFuncionario é candidata).

**Chave Estrangeira (FK):** coluna que define **como as tabelas se relacionam** umas com as outras. Nada mais é do que uma “referência” em uma tabela à PK de outra tabela.

**TUPLA / REGISTRO:** é uma **linha da tabela**, também conhecido por registro. Ela agrega um **conjunto de ATRIBUTOS**. Exemplo: uma linha da tabela pode conter vários atributos como CPF, nome, sexo, peso, etc

**RELACIONAMENTOS:** uma vez que as entidades são identificadas, deve-se então definir **como se dá o relacionamento entre elas**. São **sempre VERBOS**. Graficamente: 

Por **padrão** é **BINÁRIO**, mas **não há limites**;



- ✓ Relacionamento **PODE** ter atributo;
- ✓ Relacionamento pode ser simultâneo (“ida e volta”) ou não-simultâneo (só um caminho) Os relacionamentos simultâneos podem ser (**CAI BASTANTE**):
  - Independentes: **NÃO HÁ necessidade de avaliação simultânea de outro relacionamento** (ex: um contribuinte pode pagar sobre o lucro IRPJ e CSLL)
  - Contingentes / Paralelos (**MAIS CAI**): **impõe o estabelecimento SIMULTÂNEO DE ASSOCIAÇÕES**. Mais de um relacionamento deve ocorrer em um mesmo instante.
  - Mutuamente exclusivos: **a existência de um relacionamento EXCLUI a existência do outro** (ex: ou contribuinte tem CPF ou tem CNPJ)

A **CARDINALIDADE** dos relacionamentos pode ser:

<b>1:1</b>	Ex: <u>1</u> pessoa – só pode ter – <u>1</u> CPF
<b>1:N</b>	Ex: <u>1</u> professor – responsável por – <u>N</u> disciplinas sendo que <u>1</u> disciplina – só pode ter – <u>1</u> professor
<b>M:N</b>	Ex: <u>1</u> título - pode ser escrito - <u>vários</u> autores, sendo que <u>1</u> autor - pode escrever - <u>vários</u> títulos

Como fazer a leitura (vide exemplos abaixo):



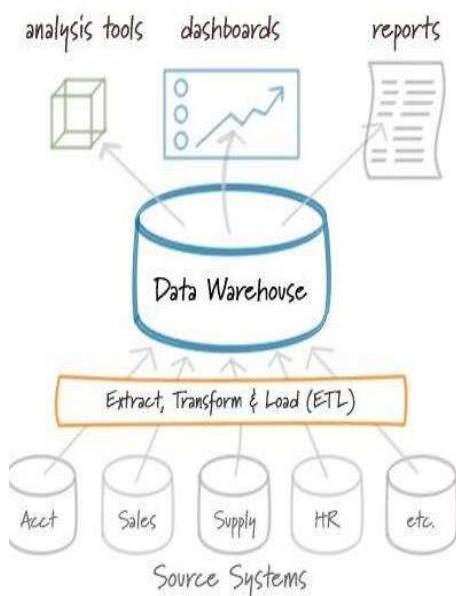
**Em azul:** 1 secretária só pode estar alocada em 1 ou 1 departamento, ou seja, não existe secretária sem departamento;  
**Em vermelho:** 1 departamento pode ter 0 ou 1 secretária, ou seja, pode existir departamento sem secretária;



- 1 cliente pode hospedar em nenhuma ou vários quartos
  - 1 quarto pode ser hospedado por nenhum ou vários clientes
- Esse é um caso de cardinalidade máxima, relação N:N, pois posso ter vários alunos cursando várias disciplinas, bem como várias disciplinas sendo cursadas por vários alunos.

## DATA MART E DATAWAREHOUSE

### DATA WAREHOUSE (DW)



Data Warehouse: **COLEÇÃO de DADOS de VÁRIAS fontes (e-mails, pesquisas, planilhas, etc.)**, internas ou externas, armazenadas sob um **esquema UNIFICADO**, em um **ÚNICO local**, que propõe sustentar a **TOMADA DE DECISÃO**. Característica do DW:

**Orientada por assunto** ("Sales", "Supply", etc.)

Variante no tempo (permite uma **visão histórica** para possibilitar a **análise temporal** e não apenas estática / momentânea)

✚ **Não volátil** (dados **sempre inseridos, nunca excluídos / alterados, SALVO correções**)

**Integrada** (várias fontes diferentes, mas que pelo ETL são padronizadas. Ex: se entrada tem *cm, mm, pol, m*, tudo deve ser convertido – "transform" para uma unidade, p.e. *cm*).

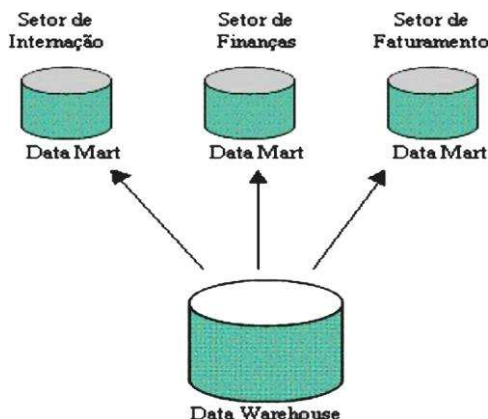
Granularidade de dados: refere-se ao nível de sumarização (resumo) do dado.

↑granularidade = ↑sumarizado = ↓nível de detalhes.

Ex: País → Estado → Cidade → Bairro. Quanto mais para a direita, maior detalhamento e menor granularidade / sumarização

### DATA MART (DM)

Data Mart (DM): nada mais é que um **SUBCONJUNTO de dados de um DW**, com **informações de interesse particular para um determinado setor**.



## ABORDAGENS PARA PROJETO DE DATAWAREHOUSE

Bottom-up: considera que um DW possa ser **composto a partir de Data Marts previamente desenvolvidos**. Tem como desvantagem uma maior dificuldade de consolidação da informação.

Top-Down: considera que se deve **desenvolver um DW completo e centralizado ANTES que partes dele**, sumarizadas, possam ser derivadas na forma de Data Marts.

### ETL – EXTRACT, TRANSFORM, LOAD

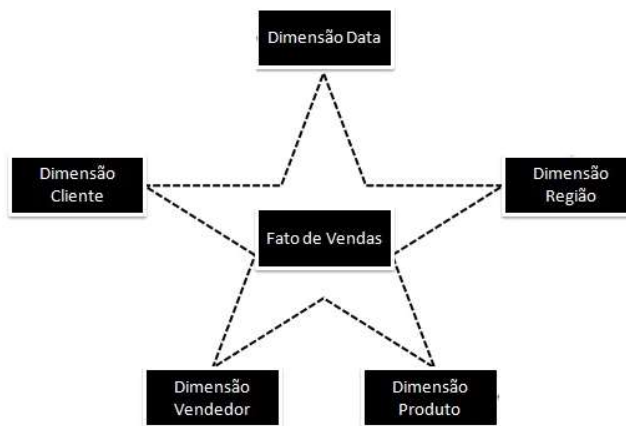
Extract: dados são **extraídos dos diversos sistemas organizacionais**, como BD, nuvem, arquivos textos, etc., ou ainda dos OLTP (Online Transaction Processing) e são **conduzidos para uma área de transição (Staging Area)** onde serão convertidos e padronizados em um único formato.

Transform: após a extração, os dados devem ser **transformados para que seja possível a carga dos dados em um DW ou DM**. Nessa fase os dados são convertidos e padronizados (por exemplo, se a fonte de dados tem informações em  $m^2$ ,  $cm^2$  e  $dm^2$  a ideia aqui é padroniza-los para um único formato, como  $m^2$ ).

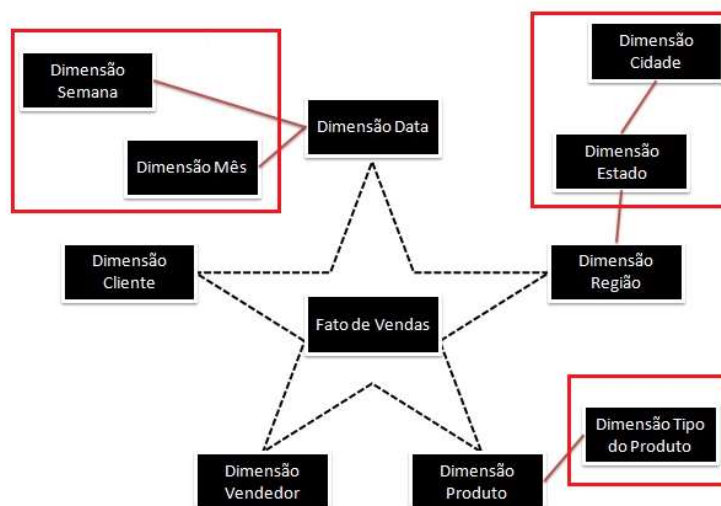
*Load*: nada mais é do que o processo de carregar os dados extraídos e transformados para um DW ou DM.

## MODELAGEM DIMENSIONAL

Esquema Estrela (star schema): há uma **Tabela de Fatos (normalmente muito grande)**, que é o elemento central, e várias Tabelas de Dimensões, sempre interligadas à Tabela de Fatos por um identificador.



Esquema Floco de Neve (snowflake): uma ou mais pontas da estrela têm conectadas a si novas tabelas de dimensões



## DATA MINING

A mineração de dados é um campo interdisciplinar que reúne técnicas de aprendizado de máquina, reconhecimento de padrões, estatísticas, banco de dados e visualização para abordar a questão da **extração de INFORMAÇÕES** a partir de **grandes bases de dados, visando a DESCOBERTA de PADRÕES, que possam ser ÚTEIS ao negócio**.

**Atenção!** Não confundir data mining com uma simples busca ("query").

## BI – BUSINESS INTELLIGENCE

**BI**: PROCESSO de tomada de decisões, no qual a info é repassada de uma forma adequada aos decisores, dando **subsídios a uma decisão embasada em info relevantes** (geralmente dados consolidadoa, de um DW ou DM).

### OTP x OLAP

A principal diferença entre OLTP (On-Line Transaction Processing) e OLAP (On-Line Analytical Processing) é o objetivo.

**OLTP**: voltado a **REGISTRAR as transações em TEMPO REAL do negócio**, com inclusão e ALTERAÇÃO (inclusive exclusão) contínua de dados. Sua capacidade de apoiar a tomada de decisões é limitada. Pense no sistema que registra as vendas de um supermercado, alimentado pelas informações de vendas em tempo real de cada caixa.

**OLAP**: voltado à **ANÁLISE, ao apoio à TOMADA DE DECISÃO**. Nele, os dados são geralmente apenas adicionados (salvo correção, por exemplo), e **não há tanta necessidade de operação em tempo real**. Imagine um Diretor de uma rede de

supermercados precisando tomar decisões. Ele não conseguiria extrair muita coisa de um sistema OLTP que registre todas as vendas em tempo real, não é mesmo? Mas essas informações podem, por meio de um processo ETL, serem convertidas em um DW alimentado, por exemplo, diariamente. Um sistema OLAP deve ele deve proporcionar facilidade para a realização de consultas, com uma navegação rápida, flexível, interativa

Drill Down: aumentar o nível de detalhe da informação

Roll Up: diminuir o nível de detalhe

Slice and Dice: modifica posição de uma informação, altera linhas por colunas de forma a auxiliar o usuário

## BIG DATA

**Big Data**: alto volume de dados, de diversas fontes distintas, que são muito difíceis de capturar, armazenar, buscar e analisar com as ferramentas tradicionais de banco de dados.

- Possui dados ESTRUTURADOS e NÃO ESTRUTURADOS;
- Objetivo: propiciar informações para subsidiar a **tomada de decisões**;
- Principais características: QUANTIDADE e VELOCIDADE

Big Data = Transações + Interações + Observações

**Big Data Analytics**: é o trabalho analítico e inteligente de grandes volumes de dados, estruturados ou não, coletados, armazenados e interpretados por **softwares de altíssimo desempenho**. Algumas fontes:

- BI
- Arquivos de log
- Conteúdo de **mídias sociais**
- IoT
- Relatórios e Indicadores

Os 5 V's do Big Data:

<b>Volume</b>	<b>Variedade</b>	<b>Velocidade</b>	<b>Veracidade</b>	<b>Valor</b>
Relacionado com a quantidade de bytes processados (Gb, Tb, Pb, etc.)	Relacionado com o formato dos dados, como textos, áudios, vídeos, etc.	Rapidez de produção e tratamento desses dados ( <i>real time</i> )	Garante a confiabilidade dos dados, i.e., que sejam verdadeiros	O resultado obtido desses dados deve trazer benefícios à entidade

**Mapreduce**: em vez de depender da velocidade de leitura e escrita de dados de um único equipamento, o MapReduce usa várias máquinas para dividir o trabalho. Veja esse vídeo de exemplo:

**Hadoop**: modelo (framework) **livre** (gratuito) para **armazenamento e processamento** em larga escala de grandes conjuntos de dados, utilizando clusters de hardware, reduzindo o custo. O Hadoop “democratizou” o MapReduce.

## SEGURANÇA DA INFORMAÇÃO

### PRINCÍPIOS

<b>Disponibilidade</b>	Informação estar <b><u>sempre DISPONÍVEL</u></b> (não será apagada). Usamos a <b>redundância (backups)</b> .
<b>Integridade (confiabilidade)</b>	Informação sempre correta e <b><u>não ter sido ALTERADA por quem não for autorizado</u></b> . Comumente utiliza-se a técnica de aplicação do <i>hash</i>
<b>Confidencialidade</b>	Informação só estará <b>ACESSÍVEL para os USUÁRIOS AUTORIZADOS</b> . Preza pela <b>privacidade e SIGILO</b> dos dados.
<b>Autenticidade</b>	Garante a <b>IDENTIDADE de quem ENVIA info na comunicação</b> . Uma forma de se obter esse princípio é através dos <b>certificados digitais</b> .
<b>Irretratabilidade (não repúdio)</b>	<b>Impossibilidade de NEGAR a autoria de uma mensagem ou arquivo</b> . Uma forma de se obter isso é através da <b>assinatura digital</b>

<b>Conformidade (legalidade)</b>	Propriedade que garante que o sistema deve <b>seguir as leis e regulamentos</b> . Preza o respeito à legislação vigente.
----------------------------------	--

## CRIPTOGRAFIA

### CHAVE SIMÉTRICA E ASSIMÉTRICA

**Criptografia de chave SIMÉTRICA:** utiliza **UMA mesma chave** tanto para codificar como para decodificar. Deve-se tomar muito cuidado com a segurança do canal pelo qual será compartilhada a chave entre as partes

**Criptografia de chaves ASSIMÉTRICAS** (= **criptografia de chave pública**): utiliza **DUAS chaves distintas**: uma **PÚBLICA**, que pode ser livremente divulgada, e uma **PRIVADA**, que deve ser mantida em segredo por seu dono.

Geralmente um **emissor** codifica com a chave **PÚBLICA** da pessoa que **RECEBERÁ** a mensagem. O texto codificado **apenas poderá ser decodificado pelo DESTINATÁRIO**, pois, **somente ele tem a chave PRIVADA** relacionada à chave PÚBLICA que originou o texto cifrado. Perceba que na **assinatura digital o processo é diferente!!**

### FUNÇÃO HASH

Quando aplicado sobre uma informação, **independentemente do tamanho** que ela tenha, gera um **resultado único e de tamanho FIXO**. É utilizado para **garantir a integridade** do conteúdo. Características da função **hash**:

- Deve ser **impossível** encontrar a mensagem original a partir do **hash**;
- O hash **deve parecer aleatório**, mesmo que o algoritmo seja conhecido.
- Uma função de hash é forte se a mudança de qualquer bit na mensagem resulta em um novo hash diferente;
- Deve ser impossível encontrar duas mensagens diferentes que levam a um mesmo hash.

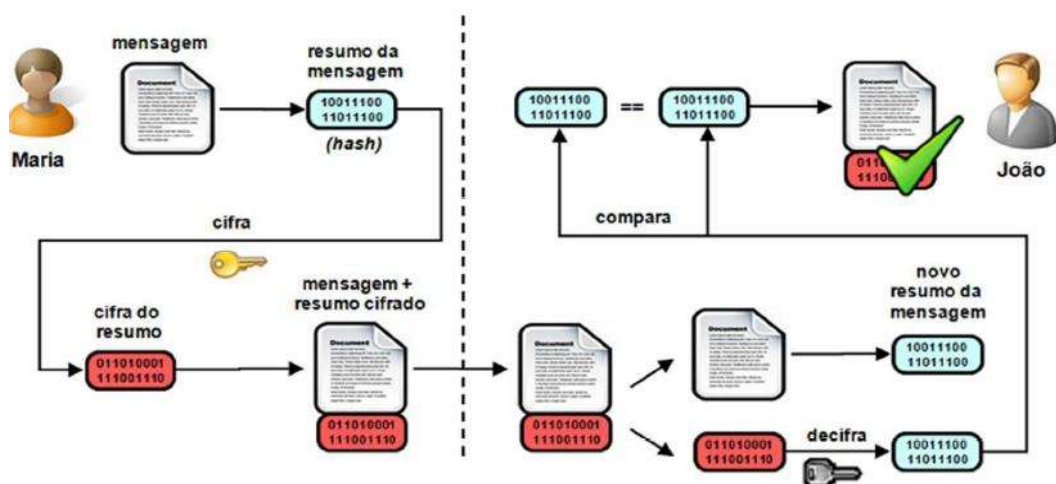
### HTTPS

**HTTPS** é um protocolo que resulta do uso do **HTTP em conjunto com o SSL**, para que as informações transferidas entre as máquinas sejam criptografadas, e quando isto ocorre, é exibido o ícone de um cadeado fechado.



### ASSINATURA DIGITAL

**Conceito:** é um método de **autenticação de informação digital**. Utiliza a **criptografia ASSIMÉTRICA** para garantir que o **destinatário confira a autenticidade** e a **integridade**. Garante também o **não-repúdio**.



#### Fluxo que garante a Autenticidade

1. Maria (**emissor**) gera o hash da mensagem;
2. **O hash** é cifrado c/ sua **chave PRIVADA** (cuidado, não é a mensagem que é cifrada!)
3. Maria envia a mensagem + hash cifrado para João.

#### Fluxo que garante a Integridade

1. João (**destinatário**) **decifra** a assinatura com a **chave PÚBLICA** de Maria, obtendo um **hash\_01**;
2. João gera, de maneira independente um novo resumo da mensagem (**hash\_02**);
3. João compara os hashes 1 e 2. Se forem iguais, a assinatura é válida, garantindo que ela não foi alterada no caminho.

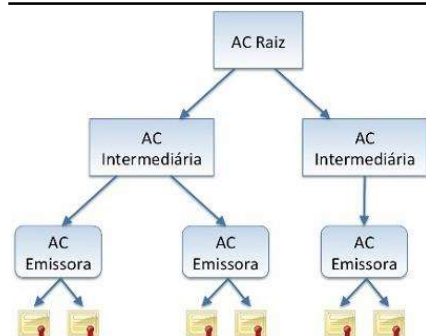
**Cuidado!** Assinatura digital não criptografa o documento (**NÃO se trata de garantia de sigilo / confidencialidade**), mas apenas utiliza da criptografia para garantir a autenticidade da mensagem.

## CERTIFICADO DIGITAL

Conceito: são **ARQUIVOS eletrônicos que servem como IDENTIDADES virtuais** (nome do portador, validade, quem emitiu, etc.) Ele tem a função precípua de **associar uma pessoa ou entidade a uma chave PÚBLICA**, i.e., responder à questão: como confiar em uma chave PÚBLICA?

O certificado digital garante que determinada entidade, organização ou pessoa é **detentor de um par de chaves criptográficas** (pública e privada = criptografia **assimétrica**), sendo a **PÚBLICA contida no próprio certificado**.

## ICP BRASIL - INFRAESTRUTURA DE CHAVES PÚBLICAS



Quem emite essa identidade? O certificado digital é emitido por uma Autoridade Certificadora (AC), uma espécie de “cartório digital”, que **também é responsável** por publicar informações sobre certificados que não são mais confiáveis.

ICPBr emite duas categorias de chaves (A e S), sendo:

A1/S1: chave **privada gerada por software** e **armazenada no HD do PC**;

A2/S2: chave **privada gerada por software** e **armazenada em token / smartcard**;

A3/S3 e A4/S4: chave **gerada por HARDware** e **armazenada em smartcard / token**, sendo que a diferença está na quantidade de bits de cada uma.

A diferença dos tipos A e S é que A é utilizada para certificação na web, e-mail, transações eletrônicas, etc., já a S certificados em com o objetivo de prover sigilo ou confidencialidade.

## ALGUMAS PRAGAS BASTANTE COBRADAS

**Ransomware:** sequestra os arquivos e programas de um computador, através da criptografia, liberando-os mediante pagamento de resgate.

**Worm:** programa malicioso que **se utiliza de uma rede para se espalhar** por vários computadores **sem que nenhum usuário interfira** neste processo (aí está a diferença entre vírus e worm = ele **não precisa de um arquivo hospedeiro**);

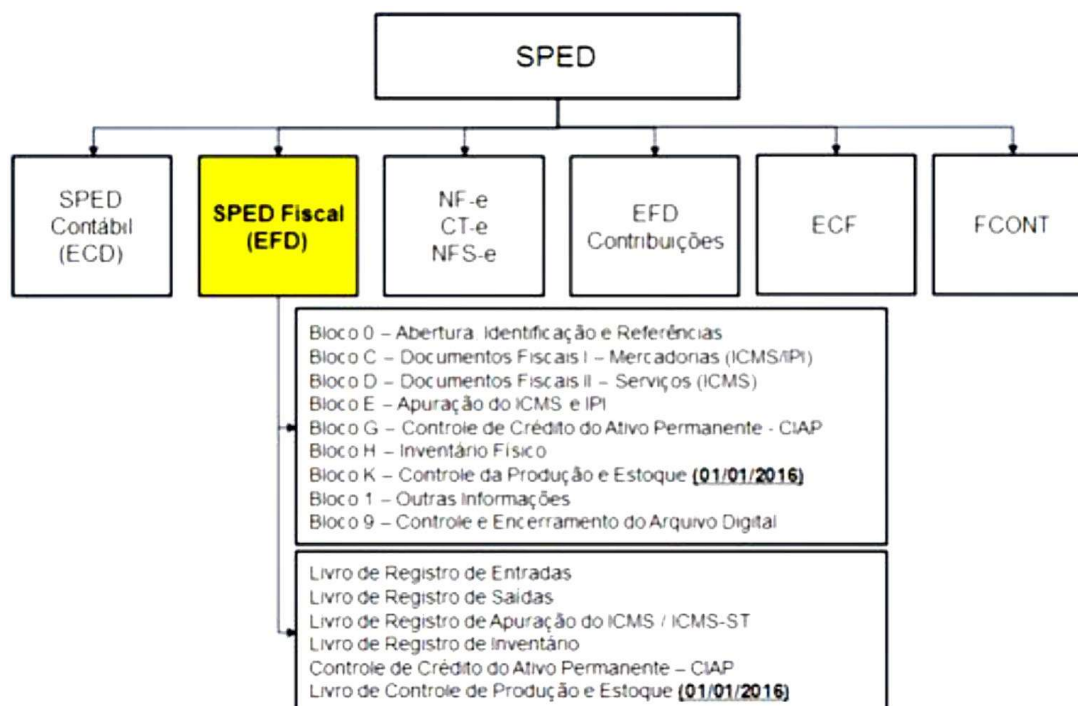
**Vírus:** quase sempre é necessário que um vírus seja **acionado através de uma ação do usuário**. Replicam-se automaticamente através de hospedeiros (arquivos).

**Trojan:** instala componentes dos quais não temos conhecimento, forçadamente (nem sempre é prejudicial, EX: quem nunca sofreu com o maldito Baidu).

## ESCRITURAÇÃO FISCAL DIGITAL – EFD

Tem como **objetivos**:

- Promover a **INTEGRAÇÃO dos fiscos**;
- Racionalizar e **uniformizar** as obrigações acessórias para os contribuintes;
- Tornar mais célere a identificação de ilícitos tributários;
- O objetivo **NÃO é punir, mas educar**;



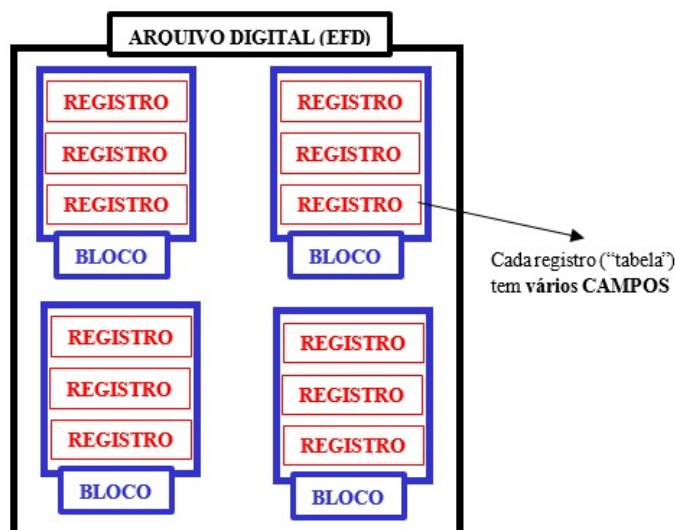
A EFD ICMS/IPI

- Uso **OBRIGATÓRIO para regime normal** (isento, imune ou não) | **NÃO ENTREGA: Simples**
- O contribuinte **deve manter todos os documentos** que suportam a EFD pelo prazo da LT
- Contribuinte deve manter EFD ICMS/IPI **para CADA ESTABELECIMENTO**
- Periodicidade (cada estado tem seu): **MENSAL, até 20º dia do mês seguinte = SC**
- **Informações sob ENFOQUE DO DECLARANTE;**
- **Retificação:** é feita uma **substituição INTEGRAL do arquivo**, i.e., não se corrige apenas uma parte, e sim manda-se um novo arquivo com todas as correções
- O **ARQUIVO DIGITAL (.TXT)** é **ASSINADO DIGITALMENTE**, e o certificado digital **só tem validade jurídica se emitido pela ICP-Brasil**. Os **certificados digitais são A1/A3**. Assinado pelo **contribuinte OU representante legal**.
- **Atenção!** Existem campos que **não têm preenchimento obrigatório**, **PORÉM**, caso a informação exista, **DEVE-SE preenche-lo**, sob pena de omissão, o que poderá acarretar penalidade.

## EFD – BLOCOS E REGISTROS

**Blocos:** Entre o registro inicial (registro 0000) e o registro final (9999), o **arquivo digital** é constituído de blocos, cada qual com um **registro de abertura**, com **registros de dados** e com um **registro de encerramento**, referindo-se **cada um deles a um agrupamento de documentos e de outras informações econômico-fiscais**.

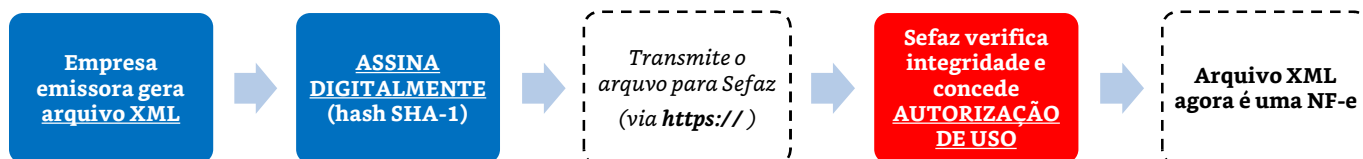
**Registros (“tabelas”):** os registros são compostos de **campos que devem ser apresentados de forma sequencial e conforme estabelecido no leiaute do respectivo registro** com todos os campos previstos **independentemente de haver ou não informação** a ser prestada naquele campo (a exclusão de campos ocasiona erro na estrutura do registro). Dentro da **hierarquia**, a **ordem de apresentação dos registros é sequencial e ascendente**. Todos os registros com a observação de “registro obrigatório” devem constar do arquivo.



## NOTA FISCAL ELETRÔNICA – NFE

**Projeto NF-e:** modelo NACIONAL de documento fiscal eletrônico, **modelo 55** - **substitui os modelos 1 e 1A** - com **validade jurídica** é garantida por: **assinatura digital do emitente** (certificado A1 / A3 – ICP-Brasil) + **autorização de uso do Fisco**.

**NFe:** documento de **existência EXCLUSIVAMENTE DIGITAL**, emitido **ANTES do FG** e armazenado eletronicamente. Não se confunde com a **DANFE, que é mera representação gráfica** (PDF) da NFe, geralmente em **ÚNICA via**, que **contém a chave de acesso da NFe** (é aquele código de barras na DANFE)



- Para emitir NFe deve ser **credenciado PREVIAMENTE na SEF**;
- O processo de **transmissão das NFe pode ser feita por lote de até 50 NFe**;
- **Atenção!** NFe **NÃO substitui** a NF do consumidor, modelo 2
- **Autorização de Uso NÃO implica convalidação das infos tributárias da NFe**;
- Se há **irregularidade fiscal do emitente / destinatário** da NFe há **DENEGAÇÃO**
  - A chave de acesso **É GERADA!**
  - A NFe é **armazenada no banco de dados!**
- Se há **problema com a NFe, como XML corrompido**, duplicado, etc., a Autorização é **REJEITADA**
  - Nesse caso, **NÃO é gerada chave de acesso**
  - O Contribuinte **pode transmitir novamente a mesma NFe**
- Pode haver problemas técnicos no site da Sefaz o que permite a emissão da NF de contingência
- **Cancelamento de NFe: até 24 horas da autorização de uso** (após isso, só via estorno);
- **Carta de Correção (CCE):** serve para sanar erro de campo específico da NFe. Alguns dados não podem ser corrigidos, como BC, alíquota, data de emissão, mudança de remetente, etc.
- **Consulta à NFe:** feita pela **Chave de Acesso** (Danfe), **disponibilizada por no MÍNIMO 180 dias**;
- **Eventos da NFe:** tudo que ocorre **após emissão da Autorização de Uso**, como **cancelamento, CCE, registro de passagem eletrônico, ciência da emissão, manifestação do fisco, NFe referenciada em CTe**.
- **DANFE:** tem como **objetivos** → acompanhar transporte; **colher firma do destinatário (comprova entrega); auxiliar escrituração de NFe pelo destinatário quando o remetente não emita NFe**.
  - Pode ser impresso em **qualquer papel, exceto jornal**.