



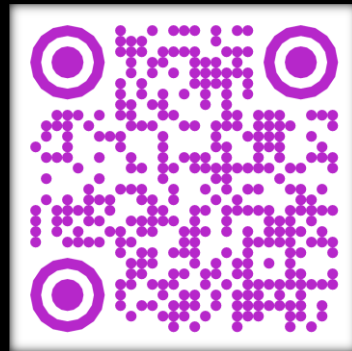
# Digital College

ENSINO DE HABILIDADES DIGITAIS

# Formação em Data Analytics

Aula 01

Professor: **MSc. Alex Souza**



# Sobre vocês

- Nome
- Formação
- Profissão
- Experiência com dados?
- Qual seu objetivo com o curso?

Formulário de alinhamento:



# Sobre o curso (Unidades)

Unidade 01 – Banco de Dados

Unidade 02 – ETL

Unidade 03 – Python para Análise de Dados

Unidade 04 – Power BI Desktop



# Sobre o curso (Unidades e Módulos)

Unidade 01 – Banco de Dados

- 1 – Banco de Dados Relacionais
- 2 – Banco de Dados Dimensionais
- 3 – SQL Avançado
- 4 – Banco de Dados - Não Relacionais



# Sobre o curso (Unidades e Módulos)

Unidade 02 – Extração, Transformação e Carga - ETL

- 1 – Processo de Descoberta de Conhecimento
- 2 – Business Intelligence
- 3 – Extração, Transformação e Carga de Dados
- 4 – Pipeline de Dados



# Sobre o curso (Unidades e Módulos)

Unidade 03 – Python para Análise de Dados

- 1 – Iniciando com Python
- 2 – Manipulação de Dados com Python
- 3 – Pandas para Análise de Dados
- 4 – Gráficos com Python



# Sobre o curso (Unidades e Módulos)

Unidade 04 – Power BI Desktop

1 – Conhecendo o Power BI

2 – Modelagem, Relacionamentos e DAX

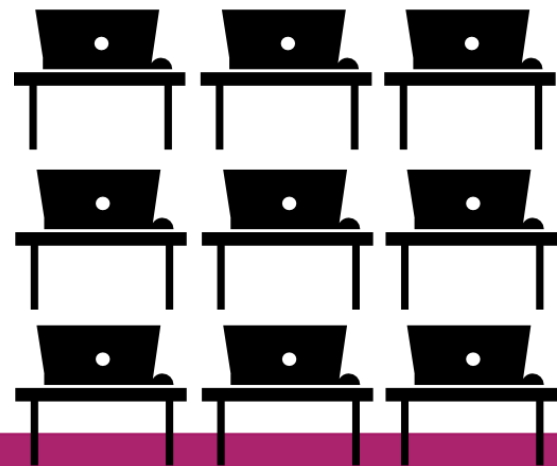
3 – Visualização de Dados





# Dicas iniciais

- Não armazenem arquivos localmente, utilizem alguma plataforma on-line (google drive, onedrive...)
- Evite utilizar pen drive ou HD externo
- Não é permitido consumo de comida na sala
- Ao sair fechem todos os programas utilizados e baixe a tampa do notebook



# Ferramentas e Tecnologias

- Arquivos – xls, txt, json, pdf, xml, etc..
- BrModelo Web ([link](#))
- Banco de Dados – PostgreSQL - PgAdmin
- Python Anaconda – Jupyter Notebook e VS Code
- Power BI
- Github, Trello, Kaggle, etc..
- Ferramentas Case



# Github



Todo material de aula, scripts e dicas ficarão em um repositório do github

- Criar uma conta no github
- Me enviar o nome do usuário (para eu compartilhar o repositório)

Repositório

- [https://github.com/aasouzaconsult/DC\\_Data-Analytics\\_01](https://github.com/aasouzaconsult/DC_Data-Analytics_01)



# Contato

O que preferem?

- Discord 
- Whatsapp 





# Unidade 1 – Banco de Dados



# Habilidades Comportamentais

- Pensamento crítico
- Resolução de problemas
- Comunicação efetiva
- Escuta Ativa
- Habilidade de apresentação
- Networking e habilidade para trabalhar em equipe
- Inteligência Emocional e Empatia
- Aprendizado contínuo (*lifelong learner*)
- Adaptabilidade (ferramenta, negócio...)



# Módulo 1: Bancos de Dados Relacionais





# Excel para Análise de Dados

Com o Excel, é possível importar, manipular, analisar e visualizar dados de diferentes fontes, como bancos de dados, arquivos de texto, arquivos CSV e outros.

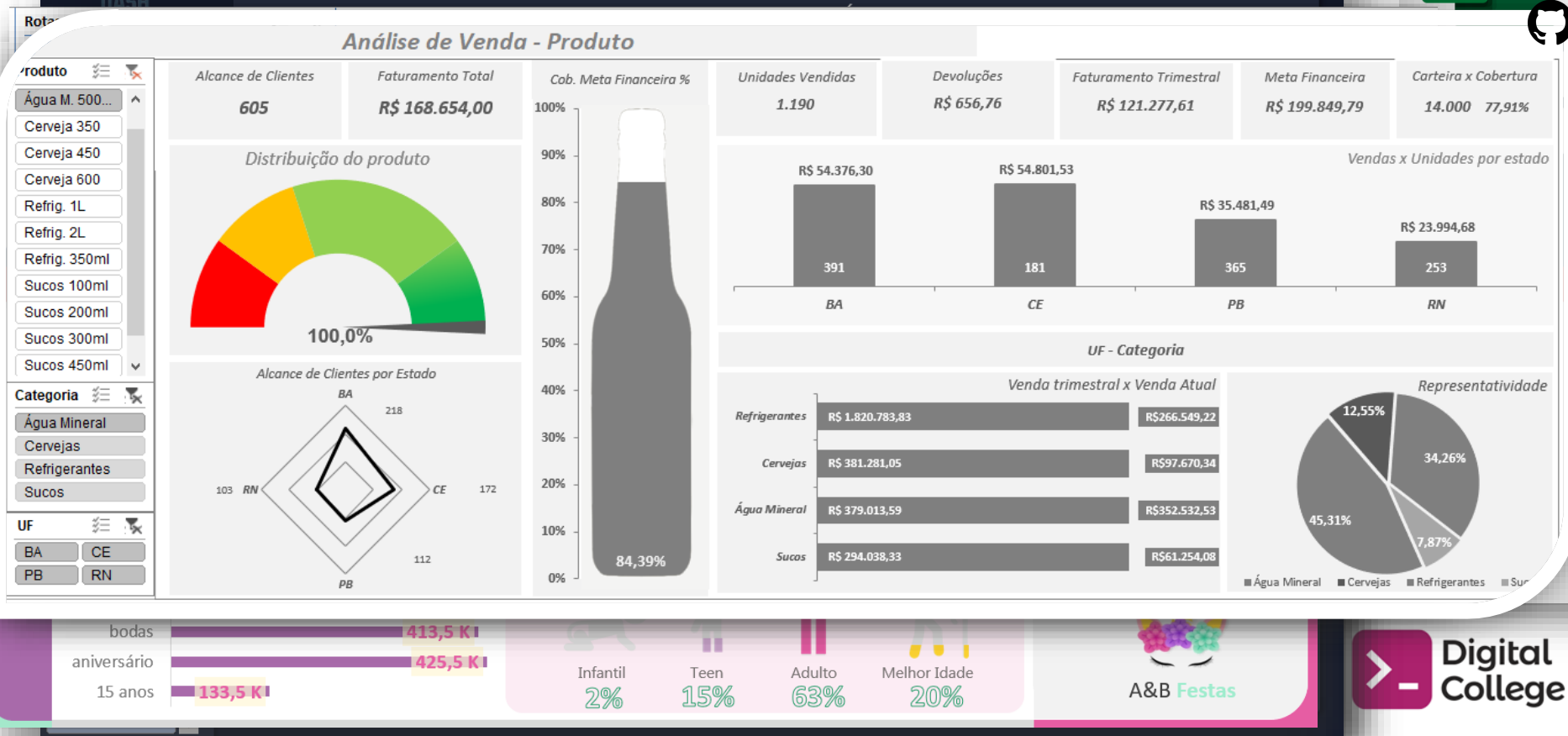
Algumas das principais funções do Excel para análise de dados incluem:

- Classificação e filtragem de dados
- Criação de tabelas dinâmicas e gráficos para visualização de dados
- Cálculos de estatísticas descritivas, como média, mediana, desvio padrão, entre outros
- Análise de regressão e correlação
- Uso de fórmulas e funções para cálculos complexos
- Importação e exportação de dados em diferentes formatos

O Excel é uma ferramenta poderosa para análise de dados, especialmente para aqueles que estão iniciando na área e ainda não estão familiarizados com ferramentas mais avançadas. Com sua interface intuitiva e grande variedade de recursos, o Excel pode ajudar a simplificar a análise de dados e fornecer insights valiosos para tomadas de decisão informadas.



# Excel para Análise de Dados - Exemplos



# O que é um Banco de Dados Relacional?

*“Um **banco de dados relacional** é um tipo de **banco de dados** que armazena e fornece acesso a pontos de **dados** relacionados entre si. **Bancos de dados relacionais** são baseados no modelo **relacional**, uma maneira intuitiva e direta de representar **dados** em tabelas.” (fonte: Oracle)*



# O que são tabela?

*“**Tabelas** são objetos de banco de dados que contêm todos os dados em um banco de dados. Nas tabelas, os dados são organizados de maneira lógica em um formato de **linha-e-coluna** semelhante ao de uma planilha. Cada linha representa um registro exclusivo e cada coluna representa um campo no registro.” (fonte: Microsoft)*



# Onde surgiram os BD? Porque?

- Foi proposto por **Edgar Codd** em 1970
- **Objetivo** de solucionar problemas com armazenamento em arquivos
  - Dificuldade de acesso aos dados
  - Isolamento dos dados
  - Anomalias de acessos concorrentes
  - Problemas com segurança
  - Problemas de integridade
- Vieram os **SGBD's** - Sistemas Gerenciadores de Banco de Dados
  - Armazena os bancos de dados / tabelas...
  - Atende todos estes pontos acima!
    - Segurança | ACID | Acessos
  - Exemplo: SQL Server, PostgreSQL, Oracle...



# ACID

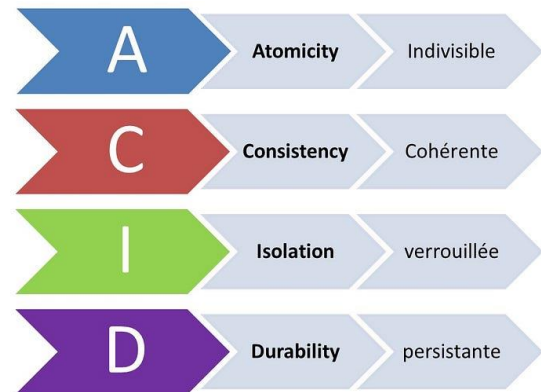
**ACID** é um conceito que se refere às quatro propriedades de transação de um sistema de banco de dados: **A**tomicidade, **C**onsistência, **I**solamento e **D**urabilidade.

**Atomicidade:** Em uma transação envolvendo duas ou mais partes de informações discretas, ou a transação será executada totalmente ou não será executada, garantindo assim que as transações sejam atômicas.

**Consistência:** A transação cria um novo estado válido dos dados ou em caso de falha retorna todos os dados ao seu estado antes que a transação foi iniciada.

**Isolamento:** Uma transação em andamento mas ainda não validada deve permanecer isolada de qualquer outra operação, ou seja, garantimos que a transação não será interferida por nenhuma outra transação concorrente.

**Durabilidade:** Dados validados são registrados pelo sistema de tal forma que mesmo no caso de uma falha e/ou reinício do sistema, os dados estão disponíveis em seu estado correto.



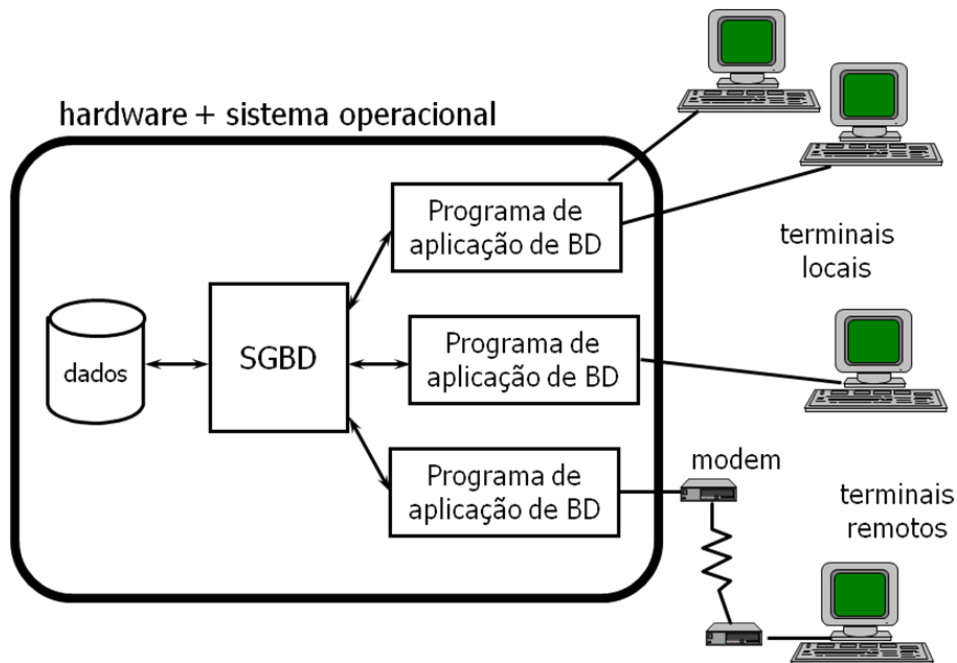
# Abstração dos Dados

- **Nível físico** – o nível mais baixo de abstração descreve *como* os dados estão realmente armazenados. Num nível físico, complexas estruturas de dados de baixo nível são descritas em detalhe.
- **Nível conceitual** – este nível descreve *quais* dados estão armazenados de fato no banco de dados e as relações que existem entre eles. O nível conceitual é usado por administradores do banco de dados, que podem decidir quais informações devem ser mantidas no banco de dados. (Analistas de Dados)
- **Nível de visões** – a maioria dos usuários do sistema de banco de dados não está interessada em todas as informações existentes no banco de dados. Cada grupo de usuários deve enxergar somente os dados que lhe dizem respeito. Assim, cada grupo de usuários tem uma visão do banco de dados. O nível mais alto de abstração é composto de visões que cada grupo de usuários tem do banco de dados. E nem interessa onde estas informações estão armazenadas.



# Evolução dos Bancos de Dados

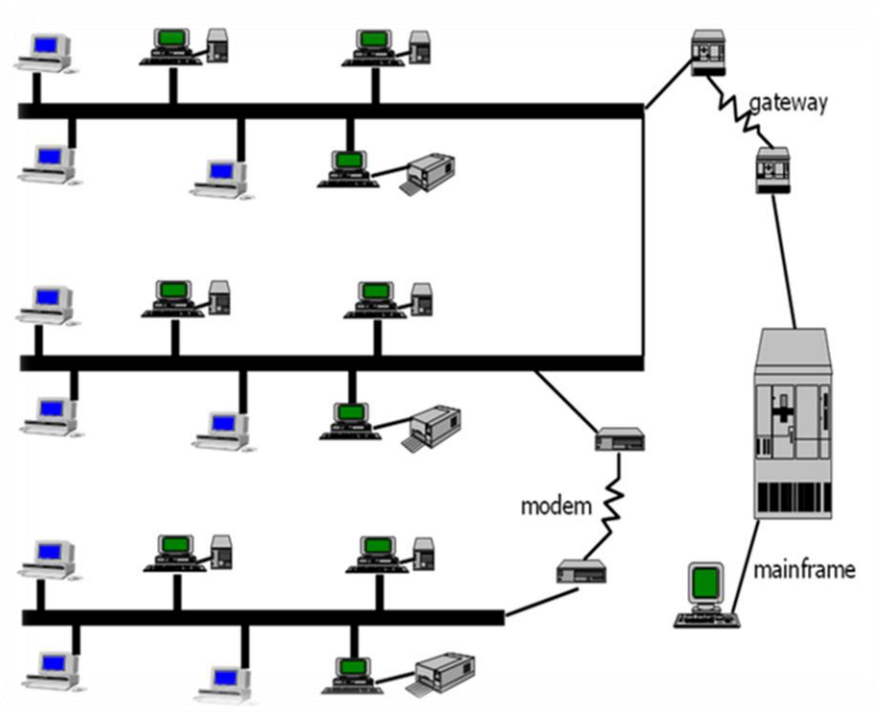
## Bancos de Dados Centralizados





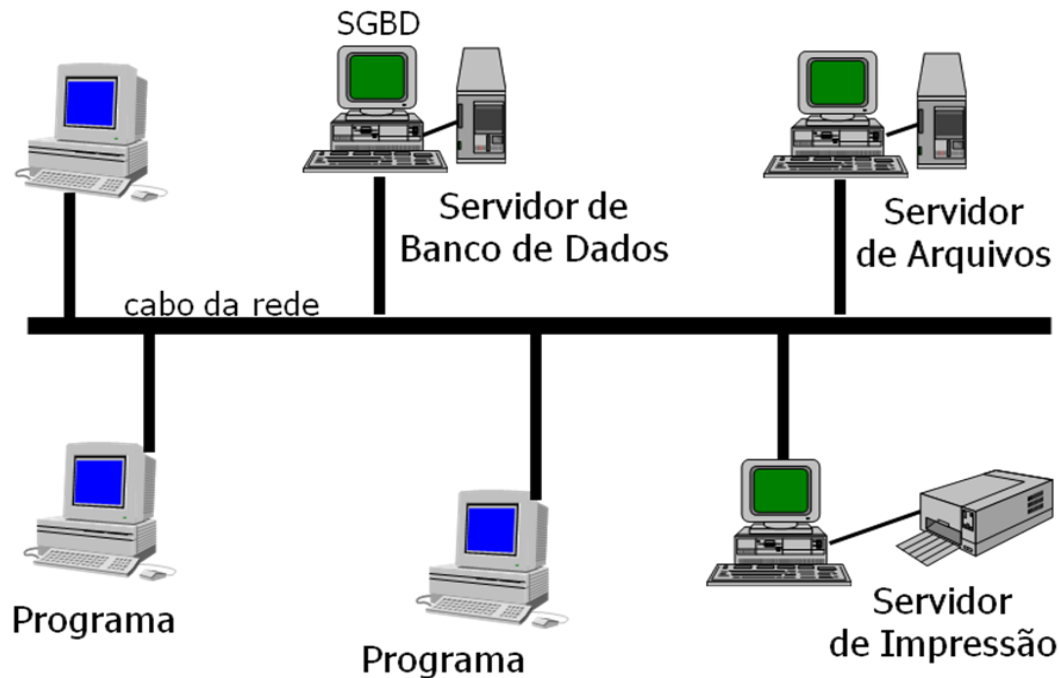
# Evolução dos Bancos de Dados

## Bancos de Dados Distribuídos



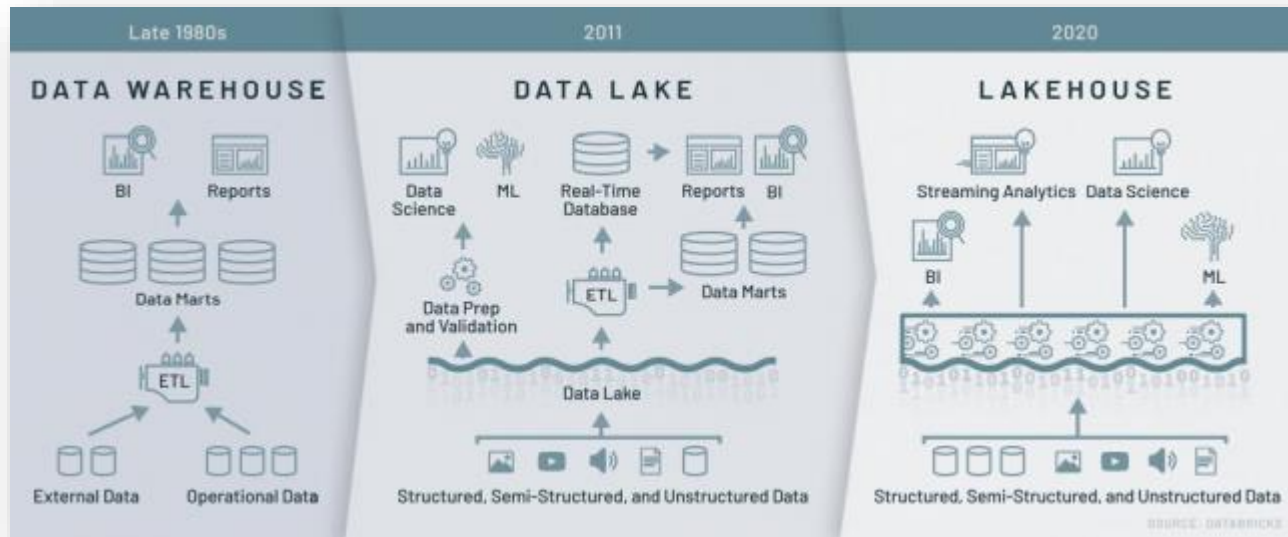
# Evolução dos Bancos de Dados

## Bancos de Dados Cliente – Servidor (com servidor de Banco de Dados)



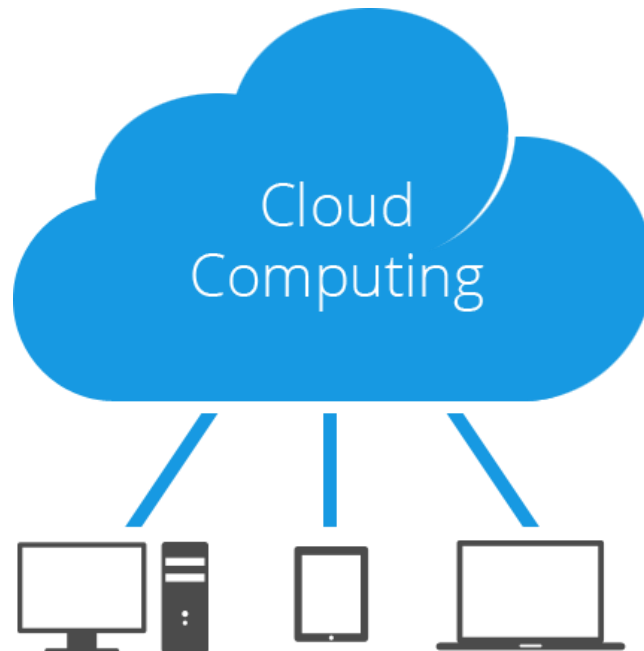
# Evolução dos Bancos de Dados

## Arquiteturas modernas



# Evolução dos Bancos de Dados

## Bancos de Dados em nuvem



# Vamos montar um banco de dados do zero?

- **Levantamento de requisitos (concepção do banco de dados)**
- **Modelo Conceitual**
- **Modelo Lógico**
- **Modelo Físico**



# Levantamento de Requisitos

*Concepção do banco de dados...*



# Requisitos do cliente

- Preciso de um banco de dados para controlar as vendas diária dos produtos da minha empresa, também precisarei saber se meus vendedores tem dependentes para poder mandar brindes em ocasiões especiais.
- Aqui trabalhamos tanto em loja própria, quanto em loja virtual.



# Colhendo requisitos




- Quais informações precisa contar na sua venda?
  - Data
  - Quantidade
  - Valor unitário e total
  - Clientes (nome, sexo, estado, idade)





# Colhendo requisitos





- Em relação aos vendedores, Eles recebem comissão?
  - Sim, recebem 
- Em relação aos dependentes, alguma informação específica deseja ter?
  - Sim, o código da Escola que Ele estuda. 
- Qual seria esse código?
  - Código do INEP 



# Colhendo requisitos



- Em relação aos produtos que vendem, alguma característica que desejam controlar?
  - Nosso produto tem uma classificação por tipo de produto, tipo A, B... 
- Mais alguma necessidade que deseja nos informar?
  - Normalmente, nossos clientes tem uma classificação com base nas compras que já fizeram, se puderem adicionar... 



# Modelo conceitual

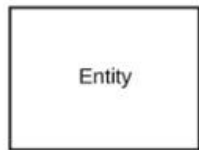
*“É o modelo de mais alto nível, ou seja, que esta mais próximo da realidade dos usuários. O nível conceitual é desenvolvido com alto nível de abstração, a partir dos requisitos do sistema, extraídos na fase de levantamento de requisitos. Esse modelo pode ser elaborado por meio de dois diagramas: **Diagrama de Entidade e Relacionamento (MER)** e/ou o Diagrama de Classes.*



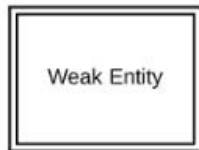
# Modelo conceitual

*Principais componentes*

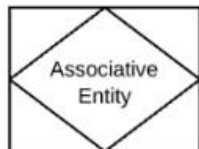




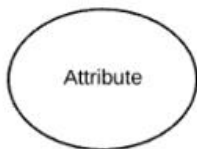
Entidade forte



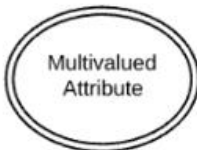
Entidade fraca



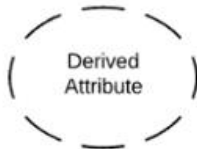
Entidade associativa



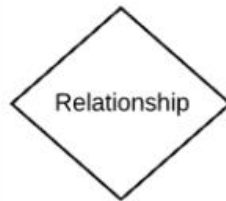
Atributo



Atributo multivalorado



Atributo derivado

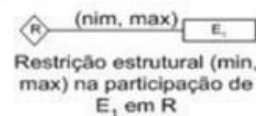
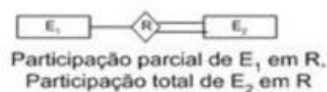
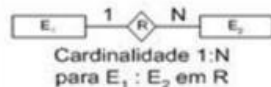


Relacionamento

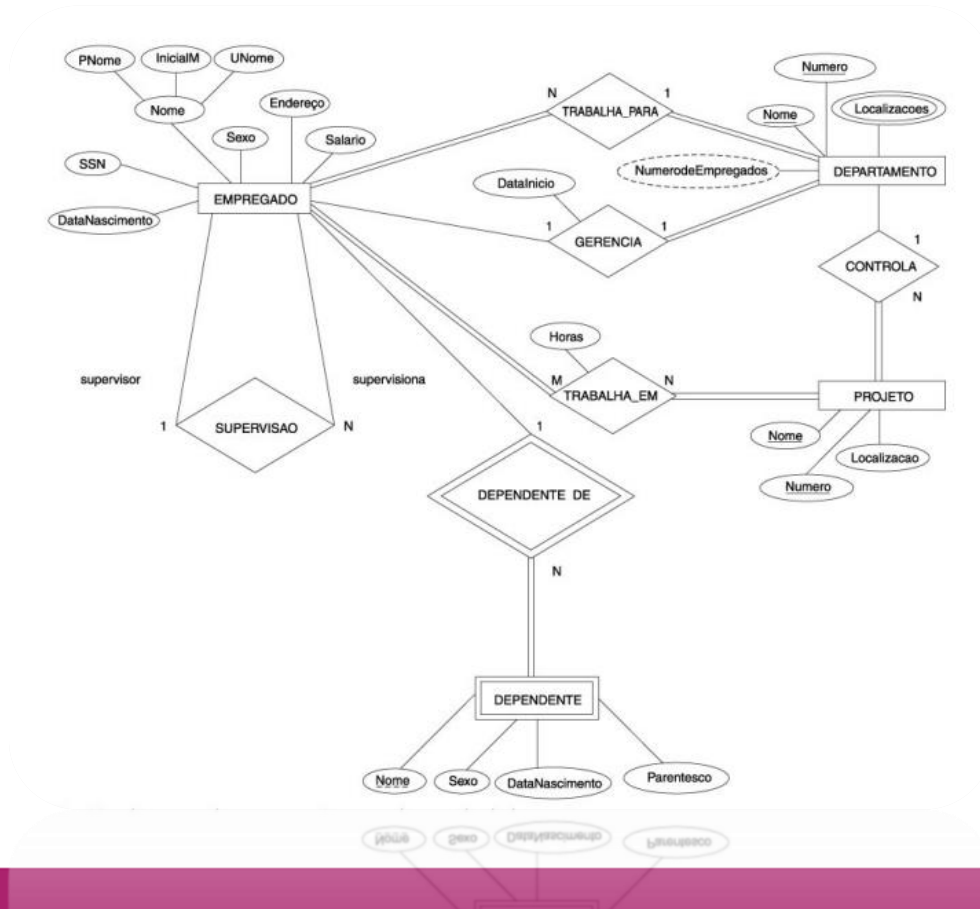


Relacionamento fraco

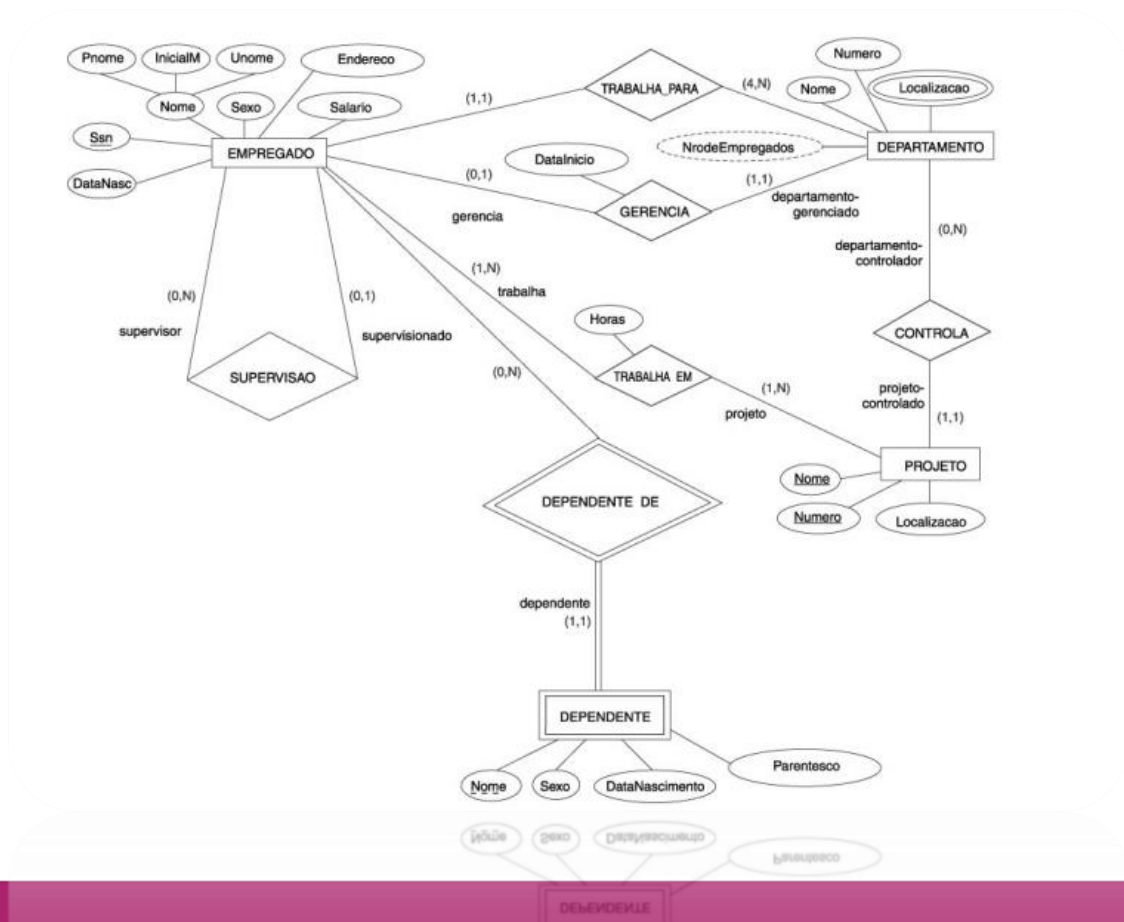
Conectividade	Peter Chan	James Martin
1:1		
1:N		
N:N		



# Exemplo 1 - MER

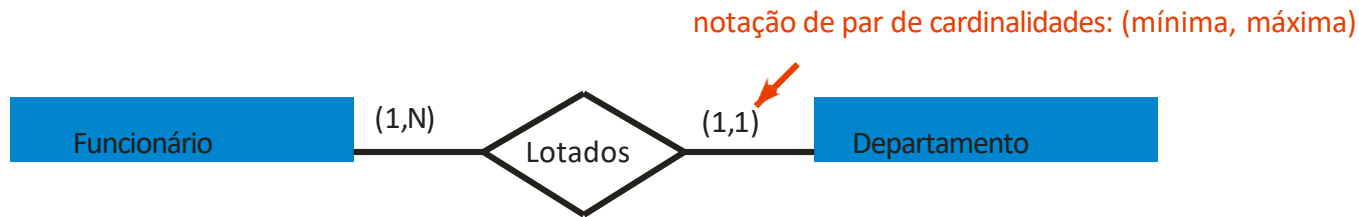


## Exemplo 2 – MER com Notação (min,max)



# Cardinalidade com Notação (min,max)

Indica se a participação das ocorrências de entidades no relacionamento é obrigatória ou opcional



*Um funcionário pode estar lotado em 1 departamento.*

*Um departamento obrigatoriamente tem que ter de 1 até N empregados lotados nele*





# Obrigad@!



Digital  
College

ENSINO DE HABILIDADES DIGITAIS

[digitalcollege.com.br](https://digitalcollege.com.br) • [@digitalcollegebr](https://www.instagram.com/digitalcollegebr)