

Dados e Repositórios Culturais

AULA 1 Introdução

Sobre o Curso

Sobre mim

**Qual o propósito deste
curso?**

Como serão as aulas?

Sobre o Curso

CO-CONSTRUÇÃO



Slides, Artigos, Materiais...



Por que falar em dados?

- A tecnologia não é apenas a digitalização da informação
- O problema da **materialidade da memória**
- O papel dos **acervos**
- A **digitalização** dos acervos e sua dimensão
- Papel da ciência da informação e multidisciplinaridade



Por que falar em dados?



- **Superfície:** Dados atuais, redes sociais, plataformas digitais.
- **Primeira Camada:** Acervos organizados, arquivos institucionais.
- **Camadas Profundas:** Memórias pessoais, arquivos informais, dados esquecidos.
- **Solo Fértil:** Saberes ancestrais, oralidades, registros não digitalizados.

Por que falar em dados?

Você já perdeu um caderno de anotações?

Um HD externo com entrevistas?

Já tentou localizar uma obra digital específica e não encontrou?



Dados x Informação x Conhecimento

Dados

- 1000 milibares
- 5,1 m/s; 95°
- 30 °C
- poucas
- 1000 mts

Informação

- Pressão atmosférica = 1000 milibares
- Velocidade e direção do vento = 5,1 m/s; 95°
- Temperatura do ar = 30 °C
- Nuvens = poucas
- Visibilidade = 1000 mts

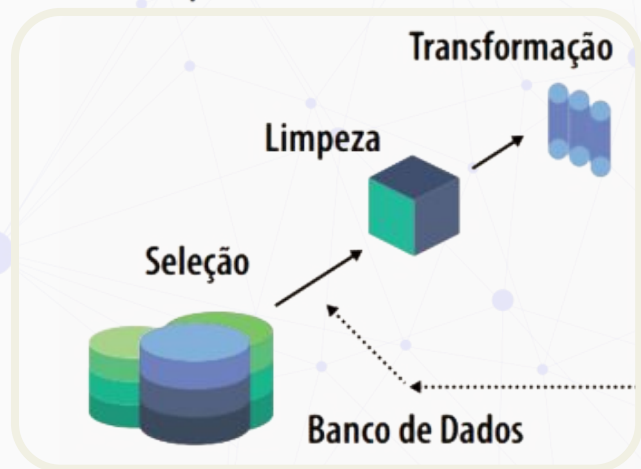
Conhecimento

- A probabilidade de chuva é baixa, portanto, posso ir à praia.

Etapas da Descoberta do Conhecimento

Mineração: Etapa Central do Processo de Descoberta de Conhecimento

Pré-processamento



Mineração

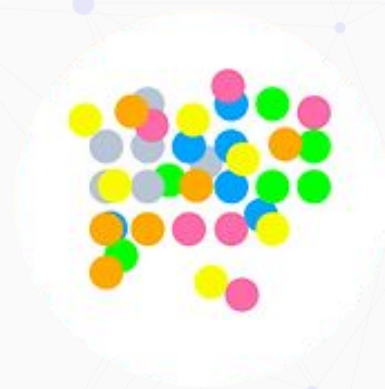
Análise do
Resultado

Conhecimento

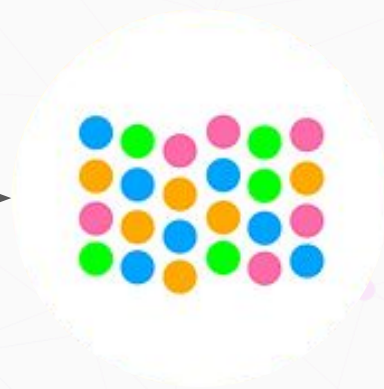
Informação

Dado

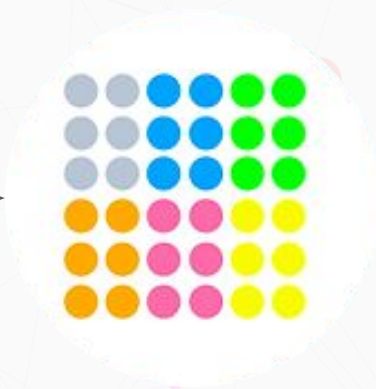
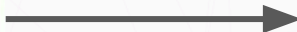
Tipos de Dados



**Não
Estruturado**



**Semi
Estruturado**

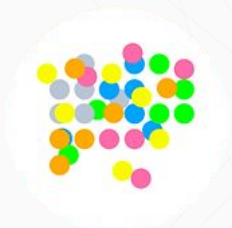


Estruturado

Não Estruturados — "O caos criativo"

São informações que não seguem nenhuma estrutura fixa. Não estão organizadas em colunas ou campos. São dados brutos, livres — mas cheios de riqueza cultural.

- Muito mais difíceis de buscar e analisar automaticamente.
- Exigem tecnologias de processamento de imagem, som, texto.
- Guardam memórias afetivas e culturais importantes.



Uma entrevista em áudio com um artista.

Vídeos de apresentações culturais.


Fotografias antigas digitalizadas.

Documentos históricos escaneados.

Semiestruturados – "Organizados, mas flexíveis"

São dados que têm alguma estrutura, mas permitem mais liberdade na forma de armazenar informações. Não estão em tabelas, mas usam marcações (tags, chaves) para dar sentido aos dados.

- Misturam ordem e flexibilidade.
- Muito usados em formatos como XML, JSON.
- Exigem ferramentas específicas para consulta.



Descrição de um acervo em um site de memória digital (em JSON ou XML).

Metadados de fotos e vídeos em repositórios culturais.

Catálogo colaborativo de um arquivo comunitário.

Estruturados — "Tudo no seu lugar"

São dados organizados em tabelas, com colunas e linhas bem definidas. Cada informação tem um campo certo. É o modo mais clássico e rígido de organizar dados.

- Fáceis de buscar, consultar e analisar.
- Usados em bancos de dados relacionais (SQL).
- Muito comuns em sistemas de cadastro e inventário.



Catálogo de obras
de um museu
(com campos
como: Título, Autor,
Ano, Técnica).

Planilha de artistas
de um festival.

Lista de livros de
uma biblioteca.

Tipos de Dados



```
{  
  "nome": "João da Silva",  
  "idade": 35,  
  "cidade": "São Paulo",  
  "telefone": "(11) 1234-5678",  
  "email": "joao.silva@email.com"  
}
```

Nome	Idade	Cidade	Telefone	Email
João da Silva	35	São Paulo	(11) 1234-5678	joao.silva@email.com

O processo **ETL (Extract, Transform and Load)** é um caminho fundamental para organizar e cuidar de dados. É um método usado para pegar informações que estão espalhadas em diferentes lugares, organizar essas informações e colocá-las dentro de um sistema – como um banco de dados ou um repositório cultural.



Extract — Extrair os Dados

Significa "puxar" os dados de onde eles estão guardados originalmente.



Podem estar:

- Em planilhas antigas.
- Em arquivos físicos que precisam ser digitalizados.
- Em sites, blogs, plataformas diversas.
- Em bancos de dados de outras instituições.

Transform — Transformar os Dados

É o momento de organizar, limpar e adaptar os dados para que possam ser usados.

Envolve:

- Corrigir erros (ex: nomes duplicados, datas erradas).
- Padronizar formatos (ex: datas todas no mesmo estilo).
- Criar novos campos ou informações (ex: adicionar categorias).
- Traduzir formatos (ex: converter um arquivo .doc para .csv).



Load – Carregar os Dados

É o momento de inserir os dados tratados em um sistema final.

Esse sistema pode ser:

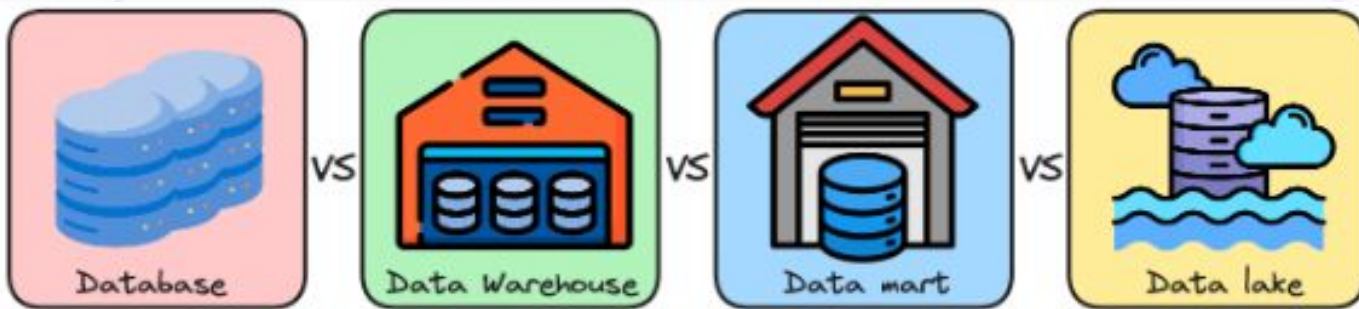
- Um banco de dados online.
- Um repositório digital (ex: Tainacan).
- Um site de memória ou acervo.
- Uma visualização pública.







Onde os Dados Vivem?

Assim como na cultura temos diferentes tipos de acervos — **bibliotecas, arquivos, museus, coleções comunitárias** — o mundo dos dados também organiza suas informações em diferentes "casas".

Cada tipo de armazenamento serve a um propósito diferente.



Onde os Dados Vivem?

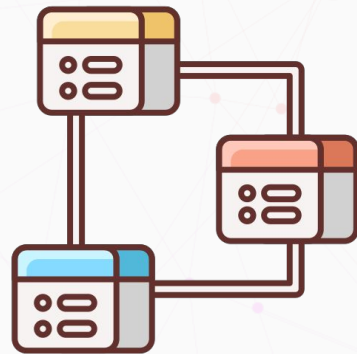
	Escopo	Finalidade Principal	Tipo dos Dados	Caso de Uso
 Database	Específico	Operação diária, dados atuais	Estruturado	Aplicações operacionais (OLTP)
 Data Warehouse	Organizacional	Análises históricas, integração de fontes	Estruturado (integrado)	Análise histórica (OLAP)
 Data Mart	Departamento	Foco em tema específico	Estruturado (segmentado)	Análise específica de negócios (OLAP)
 Data lake	Amplo	Armazenar tudo, inclusive dados brutos	Estruturado, Semi-Estruturado, Não-Estruturado	Análise exploratória, machine learning, data mining

Modelagem de Dados

Modelar dados é representar o mundo real

- O que vale a pena guardar?
- Como agrupar ou separar os elementos?
- Que relações existem entre os elementos de análise?

**Reduzir a complexidade do mundo real
para elementos essenciais.**



Modelagem de Dados

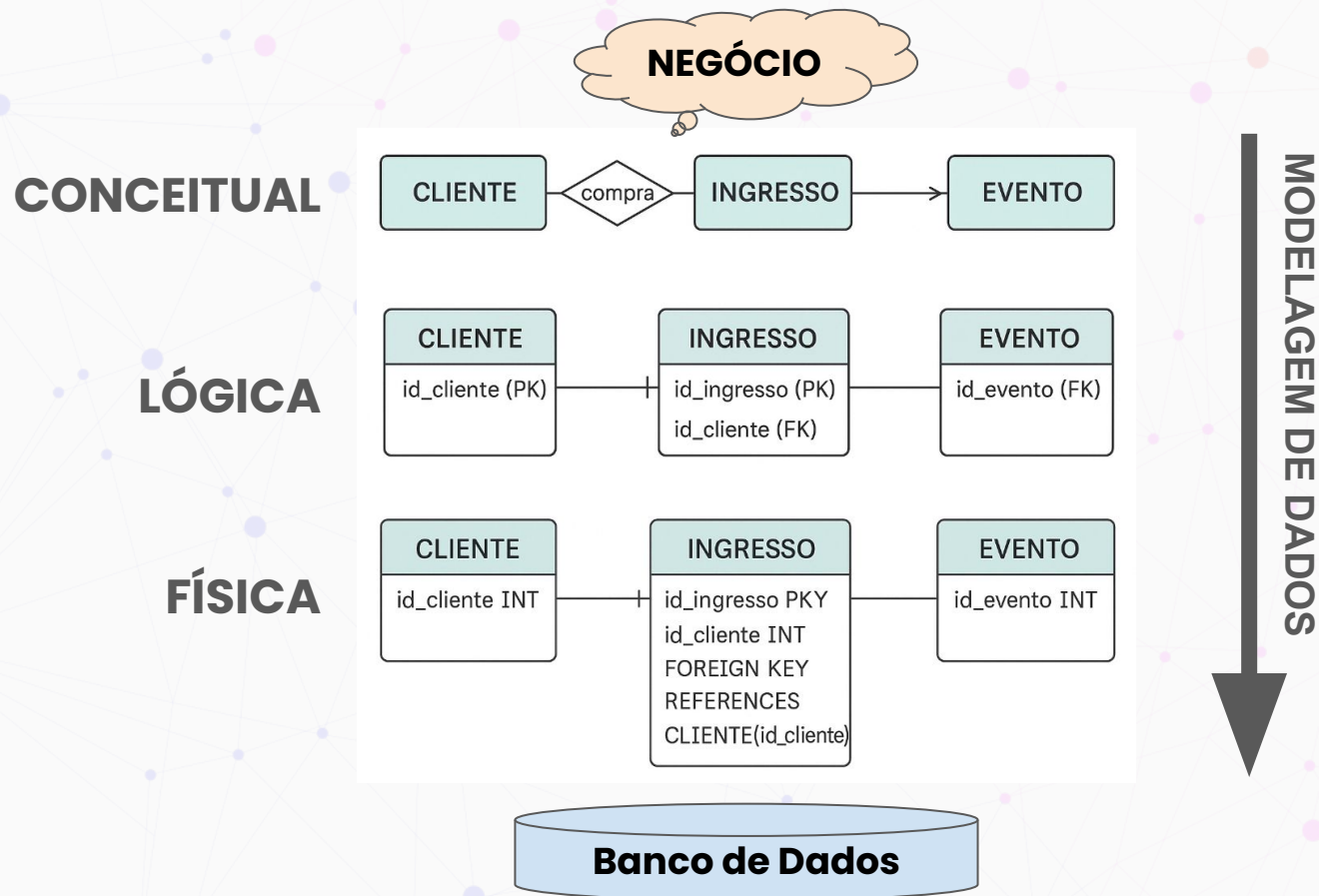
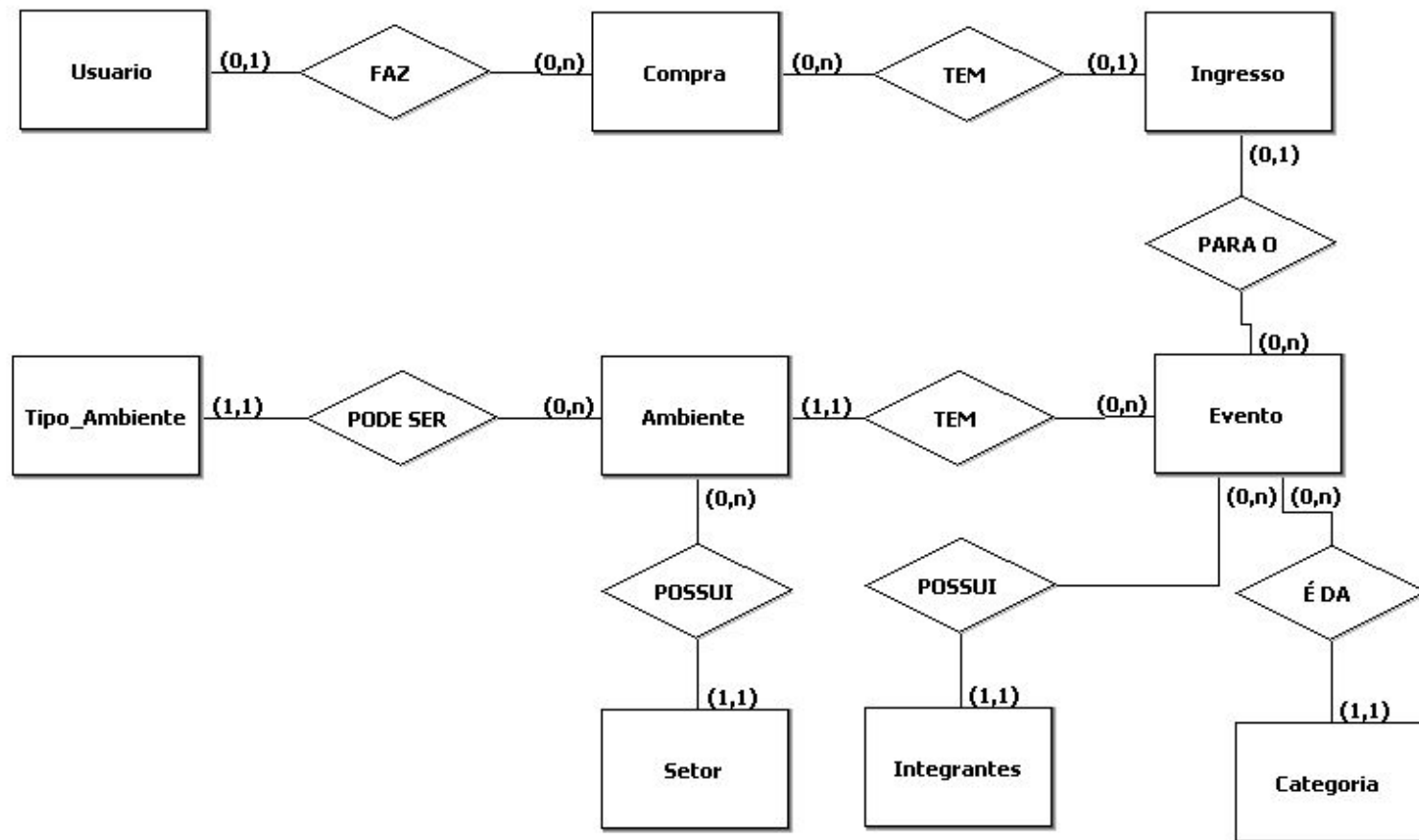


Diagrama Entidade-Relacionamento (ER)



E para consultar estes bancos de dados?



SQL



NoSQL

X



ORACLE



mongoDB



neo4j
Procedure Compiler



elasticsearch



cassandra



redis



CouchDB
relax

E para consultar estes bancos de dados?

Contexto:

Um centro cultural mantém um banco de dados com informações sobre obras artísticas de exposições passadas. Cada obra tem:

- **Título**
- **Artista**
- **Ano de criação**
- **Técnica**
- **Nome da exposição**
- **Local**

Consulta SQL

Quais obras da artista Lygia Clark participaram de exposições no Museu da Pessoa?

Tabela: obras

id	titulo	artista	ano	tecnica	exposicao_id
1	"Corpo e Memória"	Lygia Clark	1968	instalação	101
2	"Axé"	Emanoel Araujo	1982	escultura	102

Tabela: exposicoes

id	nome	local
101	"Corpo Presente"	Museu da Pessoa
102	"Identidades Visíveis"	Centro AfroDigital

Consulta SQL

Quais obras da artista Lygia Clark participaram de exposições no Museu da Pessoa?

sql

```
SELECT obras.titulo, obras.ano, exposicoes.nome AS nome_exposicao  
FROM obras  
JOIN exposicoes ON obras.exposicao_id = exposicoes.id  
WHERE obras.artista = 'Lygia Clark'  
AND exposicoes.local = 'Museu da Pessoa';
```

Consulta NoSQL (MongoDB)

Quais obras da artista Lygia Clark participaram de exposições no Museu da Pessoa?

```
json
[
  {
    "titulo": "Corpo e Memória",
    "artista": "Lygia Clark",
    "ano": 1968,
    "tecnica": "instalação",
    "exposicao": {
      "id": 101,
      "nome": "Corpo Presente",
      "local": "Museu da Pessoa"
    }
  },
  {
    "titulo": "Axé",
    "artista": "Emanoel Araújo",
    "ano": 1982,
    "tecnica": "escultura",
    "exposicao": {
      "id": 102,
      "nome": "Identidades Visíveis",
      "local": "Centro AfroDigital"
    }
  }
]
```

```
db.obras.find(
  {
    artista: "Lygia Clark",
    "exposicao.local": "Museu da Pessoa"
  },
  {
    titulo: 1,
    ano: 1,
    "exposicao.nome": 1,
    _id: 0
  }
)
```

Leitura para a próxima aula

ACERVO

REVISTA DO ARQUIVO NACIONAL

Volume 35 • Número 1 • Jul. - set. 2022



Perspectivas em
humanidades digitais



Identificando dados de pesquisa nas humanidades

Márcia Cavalcanti

CAVALCANTI, M. Identificando dados de pesquisa nas humanidades. *Acervo*, [S. l.], v. 35, n. 1, p. 1–18, 2022. Disponível em: <https://revista.an.gov.br/index.php/revistaacervo/article/view/1775>.

OBRIGADO!

Até a próxima aula!



Hora da pausa! Voltamos em:

◀◀20:00–>>