

# Megadados

Apresentação do curso, introdução a  
sistemas de gerenciamento de bancos de dados

[macielcv@insper.edu.br](mailto:macielcv@insper.edu.br)

Maciel C. Vidal

# Bem-vindos!

**Prof: Maciel C. Vidal e Cassia de O. Fernandez**

[macielcv@insper.edu.br](mailto:macielcv@insper.edu.br)

Prof auxiliar: Elisa M. Malzoni

Não enviar e-mail!

Aulas:

- Segundas e quartas, 7:30 – 9:30

Atendimento:

- Seg, 12:40 – 14:10 (presencial)

# Objetivos de aprendizado

- **Entender o que são megadados** (big data) e quais os desafios inerentes a dados com esta escala, complexidade, e requisitos de performance
- Dado um problema, **estabelecer uma estratégia de trabalho** com megadados (integração, armazenamento, processamento, tomada de decisões)
- **Projetar software analítico** capaz de utilizar estratégias de computação distribuída para tratar de forma eficaz grandes volumes de dados
- **Aplicar técnicas** de recuperação de informação e mineração de dados.
- **Descobrir e avaliar criticamente**, de forma autônoma, tecnologias emergentes em big data.

# Estrutura do curso

- Parte I: Bancos de dados relacionais
  - Modelagem
  - SQL
  - Sistemas
- Parte II: Dados em larga escala
  - NoSQL
  - Processamento em lote: MapReduce e Spark
  - Máquinas de busca e recuperação de informação

# Instrumentos de avaliação

## Projetos:

- (P1) Projeto 1: banco de dados relacional
- (P2) Projeto 2: big data
- **Nota projetos (NP):**
  - $NP = (P1 + P2) / 2$

## Avaliações:

- (AI) Avaliação intermediária
- (AF) Avaliação final
- **Nota avaliações (NA):**
  - $NA = AI*0.4 + AF*0.6$

## Atividades:

- Diversas ( $T_1, T_2, \dots, T_n$ )
- **Nota atividades (NT):**  $\min(10, \text{SUM}(T_1, T_2, \dots, T_n) / (n - 2))$ 
  - Duas atividades podem não ser entregues sem impactar na nota
  - Atividades com autograding podem ser entregues com atraso (25% da nota) até a data da AF.

## Nota final (NF):

- Se entregou e tirou pelo menos **D** em todos os projetos (P1, P2)
- E tirou pelo menos **3** em todas as Avaliações (provas)
- E nota avaliações (NA), nota projetos (NP) e nota atividades (NT) forem  $\geq 5$ :
  - $NF = NA*0.6 + NP*0.3 + NT*0.1$
- Caso contrário:  $NF = \min(AI, AF, P1, P2, NT)$

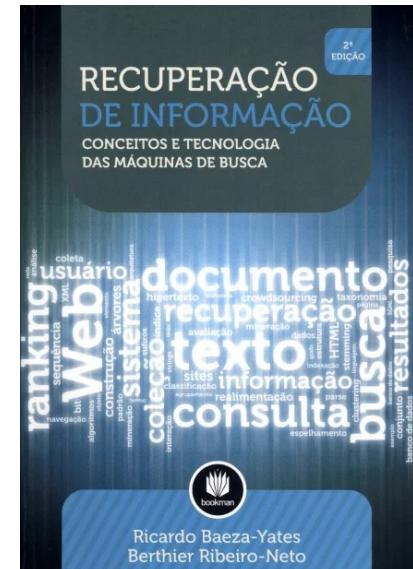
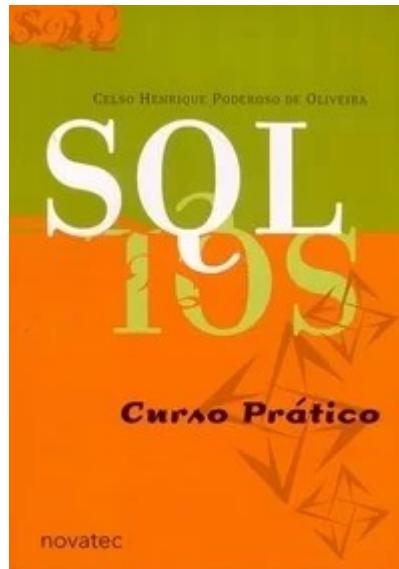
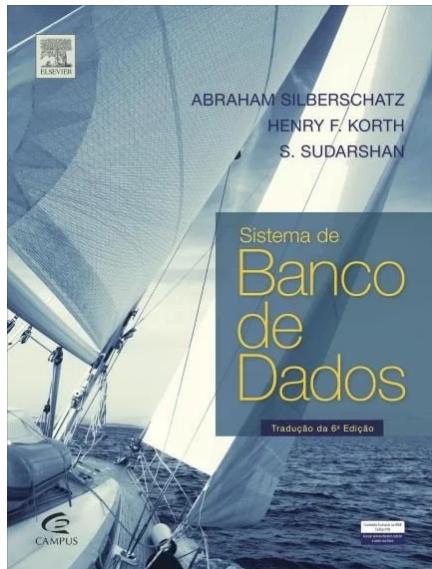
# Ferramentas de IA

- Assuma como não permitido o uso de IA.
  - Exceto quando a permissão estiver especificada.
- Cuidado com compartilhamento de conta.
  - Pode levar à reprovação e código de ética.
- Poderão existir etapas extras de verificação em todas as atividades (com ou sem permissão de IA).
  - Você precisa entender o conteúdo (vide objetivos).
  - É inadequado e proibido copiar resposta, seja da Internet, colega, livro, IA.
  - Pode levar à reprovação (mesmo com nota na atividade).

# BIBLIOGRAFIA

## BÁSICA

<b>1</b>	SILBERSCHATZ, A.; KORTH, H. F.; SUDARSHAN, S. <b>Sistema de Banco de Dados.</b> 6a Ed. Rio de Janeiro: Campus, 2012.
<b>2</b>	OLIVEIRA, C. H. P. <b>SQL: Curso Prático.</b> Novatec, 2002.
<b>3</b>	RIBEIRO-NETO, B.; BAEZA-YATES, R. <b>Recuperação de Informação - Conceitos e Tecnologia Das Máquinas de Busca.</b> 2 <sup>a</sup> Ed. 2013, Bookman



# Banco de dados: porque?

- Quais formas de armazenamento vocês conhecem?!

# Banco de dados: porque?

- Quais formas de armazenamento vocês conhecem?!
  - Um guardanapo de papel!
  - Um arquivo de texto no seu laptop!
  - Uma planilha Excel!
  - Etc...

Nem sempre precisamos de um banco de dados para armazenar dados!

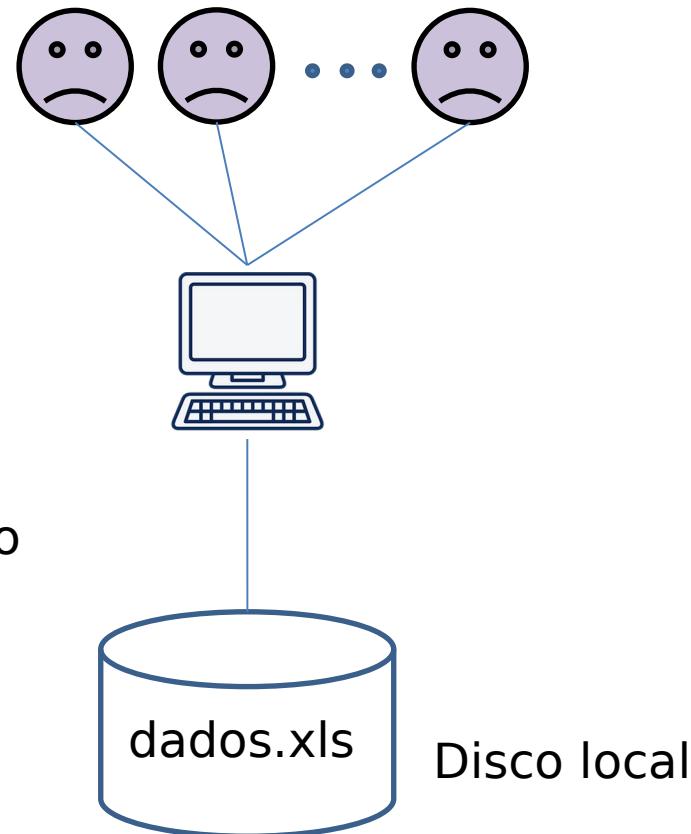
- Quando será que **precisamos** de um **banco de dados**?

# Exemplo



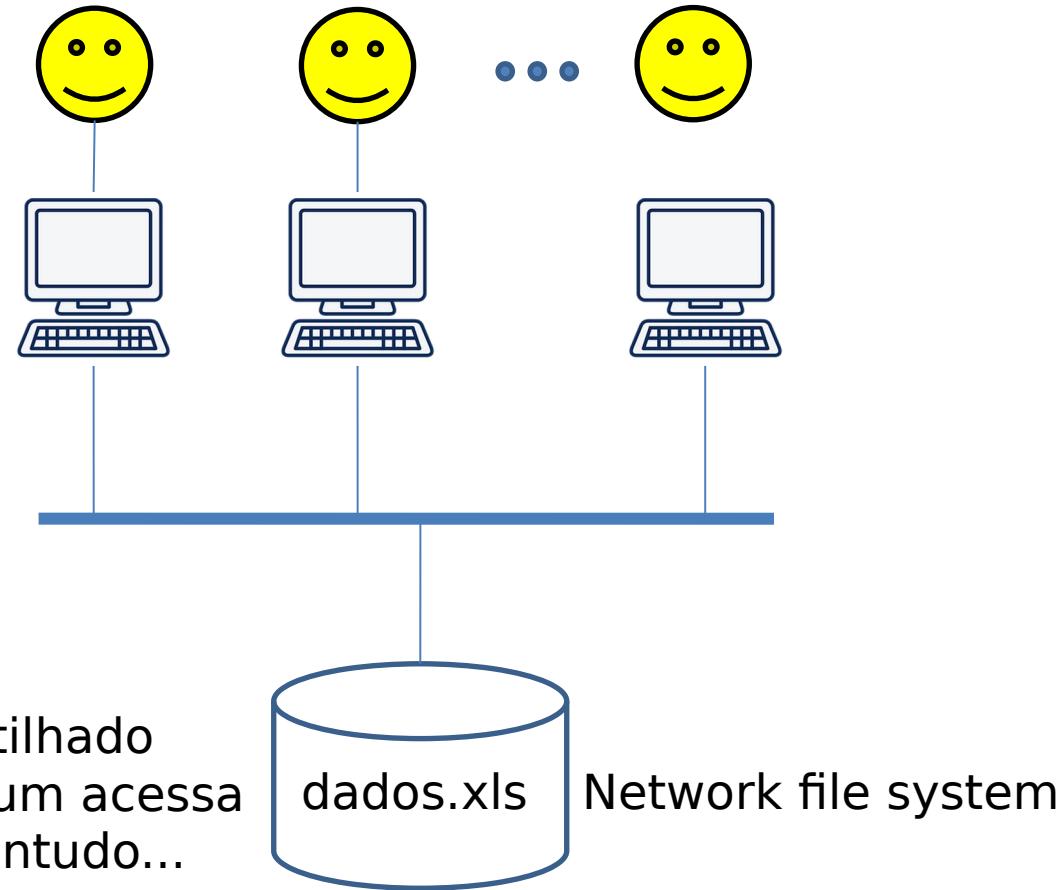
Só um usuário, arquivo pequeno...  
Não precisa de banco de dados!

# Exemplo

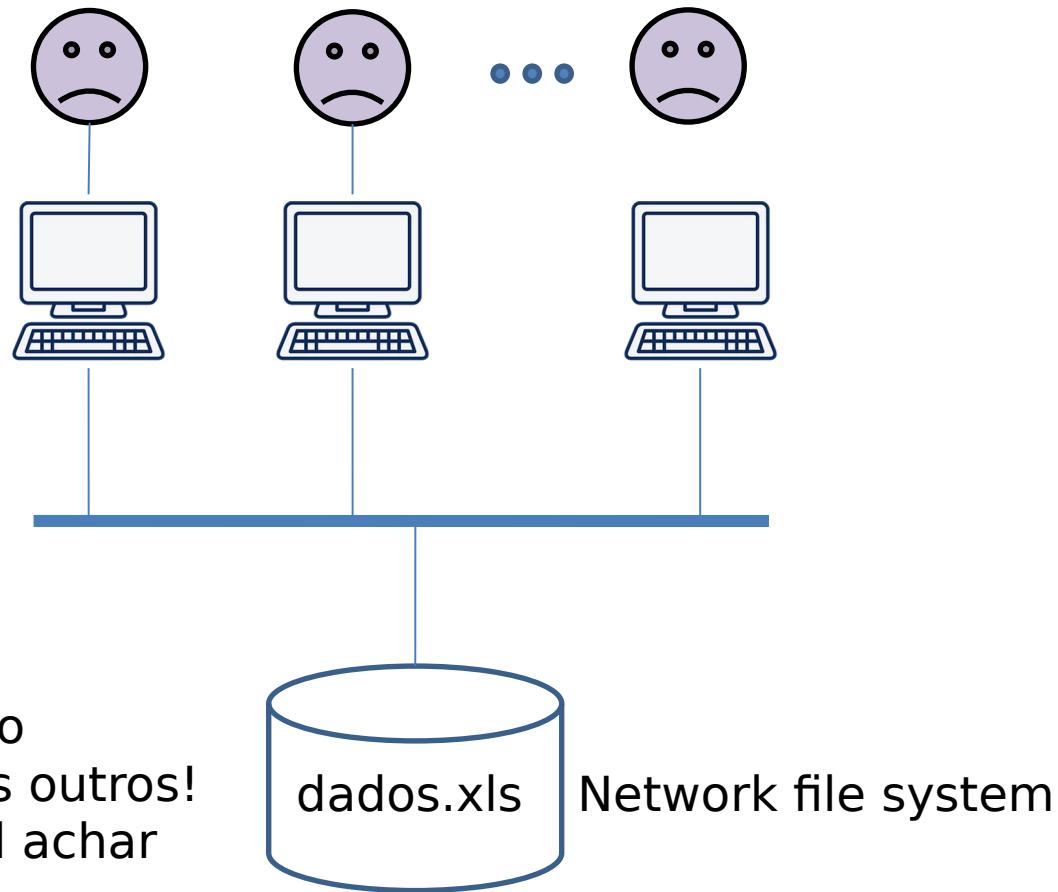


Muitos usuários formando  
fila para conseguir acessar o  
terminal!

# Exemplo



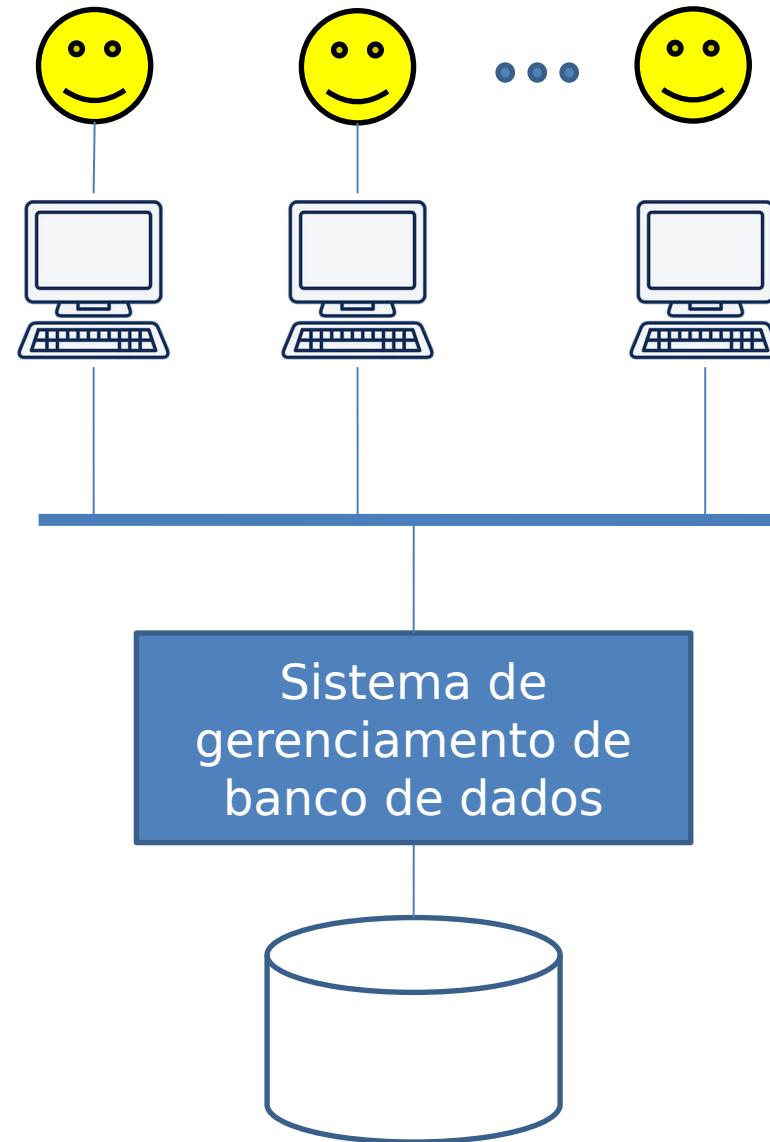
# Exemplo



- Não dá para trabalhar simultaneamente no arquivo sem estragar o trabalho dos outros!
- Dados muito grandes, difícil achar resultados
- Dados sigilosos estão expostos!

**Sistemas de gerenciamento de bancos de dados  
SGDB**

# Exemplo



# Motivos para ter um sistema de gerenciamento de banco de dados

- Tamanho
- Velocidade
- Conveniência
- Precisão
- Proteção
- Robustez

Pode não caber na RAM!

Existem maneiras eficientes de armazenar e recuperar dados!

O SGBD já vem com mecanismos sofisticados de consulta!

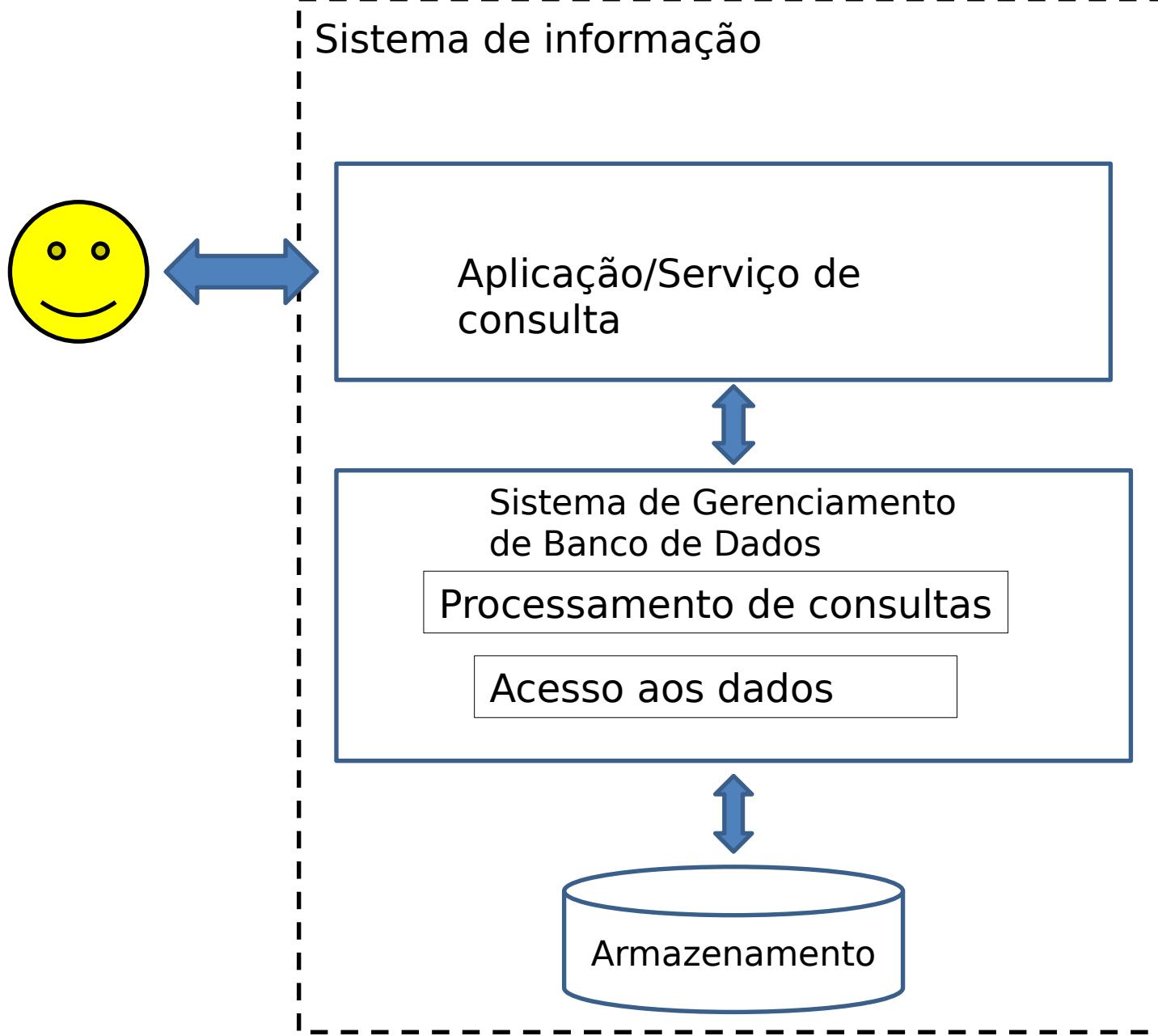
Um bom projeto evita redundâncias gerencia transações e mantém consistência!

Controle de acesso e registro de atividade!

Proteção contra falhas!

# Aplicações

- Vendas e estoque
- Recursos humanos e informações empresariais
- Dados científicos
- Informações geográficas
- Multimídia
- Jogos online
- Etc!



# Tipos de banco de dados

- **Relacional** (também chamado de bancos de dados SQL): representa os dados usando o modelo relacional, onde dados são representados através de tabelas bidimensionais.
  - Este é o modelo mais usado em bancos de dados atualmente

# Tipos de banco de dados

- **NoSQL**: bancos de dados não-relacionais, dentre os quais destacam-se:
  - Key-value stores (e.g. Redis)
  - Document stores (e.g. MongoDB)
  - Column-oriented (e.g Cassandra)

(Artigo interessante:

<http://www.dataversity.net/review-pros-cons-different-databases-relational-versus-non-relational/>

# Tipos de banco de dados

- **NewSQL:** Nova geração de bancos de dados que mesclam as vantagens de alguns tipos de bancos NoSQL (como escalabilidade e disponibilidade) com garantias de consistência transacional do SQL.
  - Exemplo: Google Spanner

(Artigo interessante:

<http://www.odbms.org/blog/2018/03/on-rdbms-nosql-and-newsql-databases-interview-with-john-ryan/>)

# Ter feito para a próxima aula

## Instalar

- Anaconda ou venv para as aulas
  - Alguma versão de Python 3 com **Jupyter Notebook**
    - Recomendo 3.12
  - Requirements do repositório de aulas
- **MySQL Community Server**
- **MySQL Workbench**
- **Autograding da disciplina**

# Agora...

- **Abra o notebook da aula para:**
  - **Instalar**
    - Anaconda ou alguma versão de Python 3 com **Jupyter Notebook**
    - **MySQL Community Server**
    - **MySQL Workbench**
  - **Autograding**
    - **Instalar pacote autograding**
    - **Configurar autograding**

# Insper

[www.insper.edu.br](http://www.insper.edu.br)