

# *Apresentação*

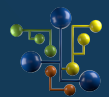
---

A Formação Analista de Dados é um programa de aperfeiçoamento profissional 100% online e 100% em português, com certificado de conclusão.

Seja Muito Bem-vindo(a)!

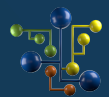
## Formação Analista de Dados





# SQL Para Data Science





# Window Functions, Subqueries e Tratamento de Datas





# Como Trabalhar com Grandes e Complexos Conjuntos de Dados?

Os dados tem alto volume, dá trabalho para carregar, demora demais e os dados tem problemas com valores nulos?

Ok, bem-vindo(a) ao mundo real!

O que fazer em casos assim (que será a maioria dos casos)?

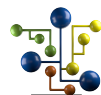
A – Sentar, chorar e reclamar.

B – Pedir ao universo que a vida seja mais fácil, o volume de dados seja menor e que não tenha problemas.

C – Desistir e dizer ao seu chefe ou cliente que não consegue fazer e que contratem outra pessoa.

✓ **D – Desenvolver uma mentalidade orientada à solução de problemas e buscar soluções.**

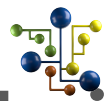




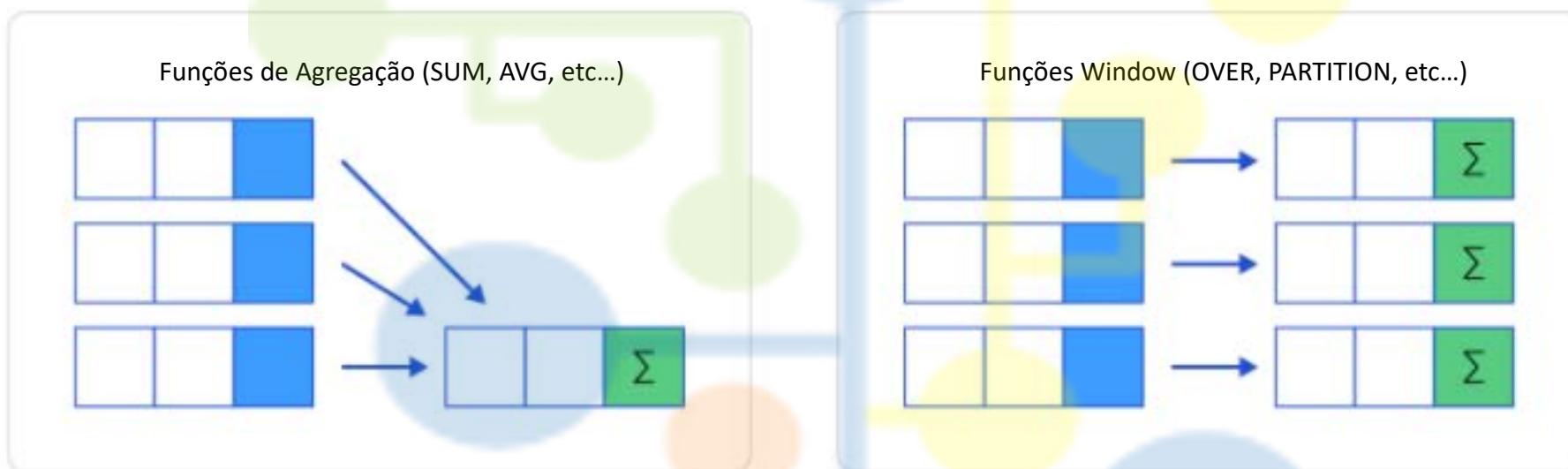
# Como Trabalhar com Grandes e Complexos Conjuntos de Dados?

- O MySQL Workbench é uma aplicação Java e apresenta diversos problemas para carregar grandes e complexos conjuntos de dados. Podemos então usar a linha de comando com o LOAD DATA (mostrado a você na solução do exercício 3).
- Se os arquivos apresentarem problemas, pode ser necessário usar um editor de texto, editar e ajustar os arquivos antes da carga.
- Podemos ainda usar uma ferramenta de *split* e dividir um grande arquivo em 3 ou 4 partes sem abrir o arquivo. E então usamos uma dessas partes para carregar os dados e construir nossas queries SQL.
- Se ainda assim a carga apresentar problemas, use outras alternativas como PostgreSQL, SQLite ou mesmo o Amazon RDS na nuvem!
- Conhecimento de sistema operacional é fundamental. Não negligencie esse conhecimento.
- Desenvolva uma mentalidade orientada à solução de problemas, busque soluções e veja a mágica acontecer na sua vida profissional. Você estará à frente de 90% dos profissionais do mercado.

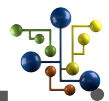




# Abrindo uma Janela Para os Dados







# Abrindo uma Janela Para os Dados

id	date	time
1	2019-07-05	22
1	2019-04-15	26
2	2019-02-06	28
1	2019-01-02	30
2	2019-08-30	20
2	2019-03-09	22

1

**PARTITION BY id**

id	date	time
1	2019-07-05	22
1	2019-04-15	26
1	2019-01-02	30

id	date	time
2	2019-02-06	28
2	2019-08-30	20
2	2019-03-09	22

2

**ORDER BY date**

id	date	time
1	2019-01-02	30
1	2019-04-15	26
1	2019-07-05	22

id	date	time
2	2019-02-06	28
2	2019-03-09	22
2	2019-08-30	20

3

**AVG(time)**

**ROWS BETWEEN 1 PRECEDING  
AND CURRENT ROW**

id	date	time	avg_time
1	2019-01-02	30	30
1	2019-04-15	26	28
1	2019-07-05	22	24
2	2019-02-06	28	28
2	2019-03-09	22	25
2	2019-08-30	20	21





# Muito Obrigado!

Continue Trilhando uma Excelente Jornada de Aprendizagem!

