

# Data Lake

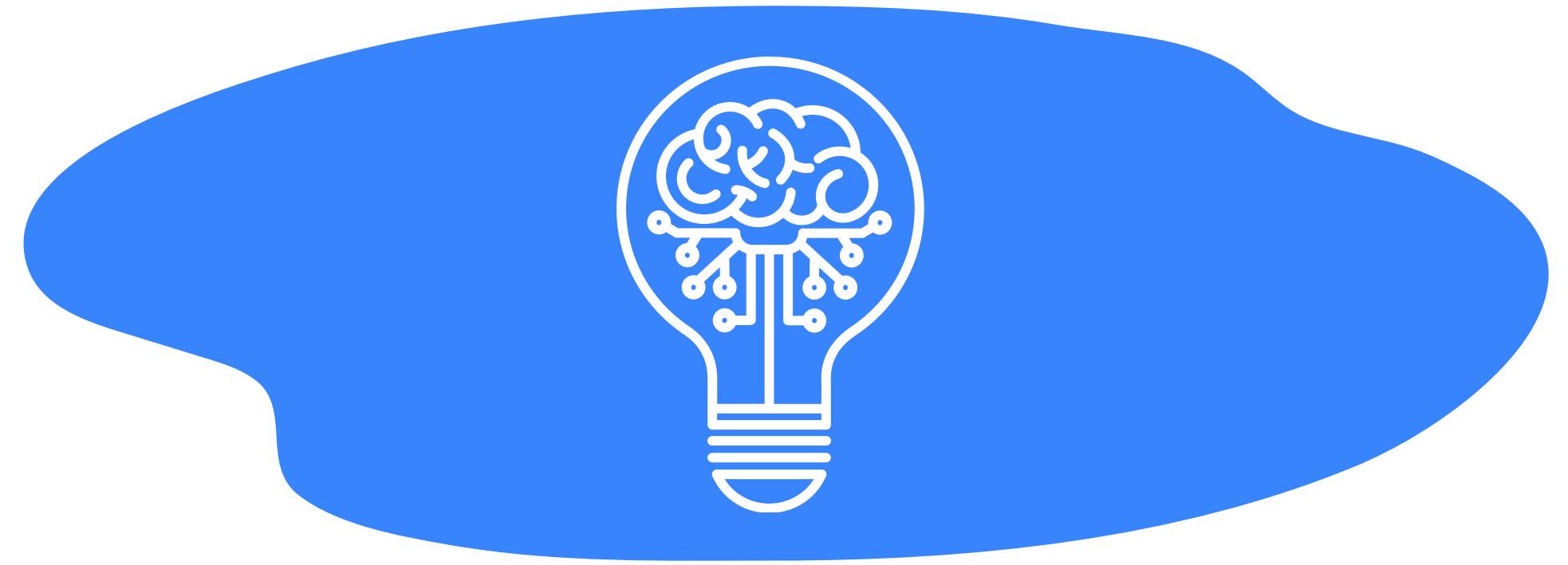
O que é?



**Renato Coelho**

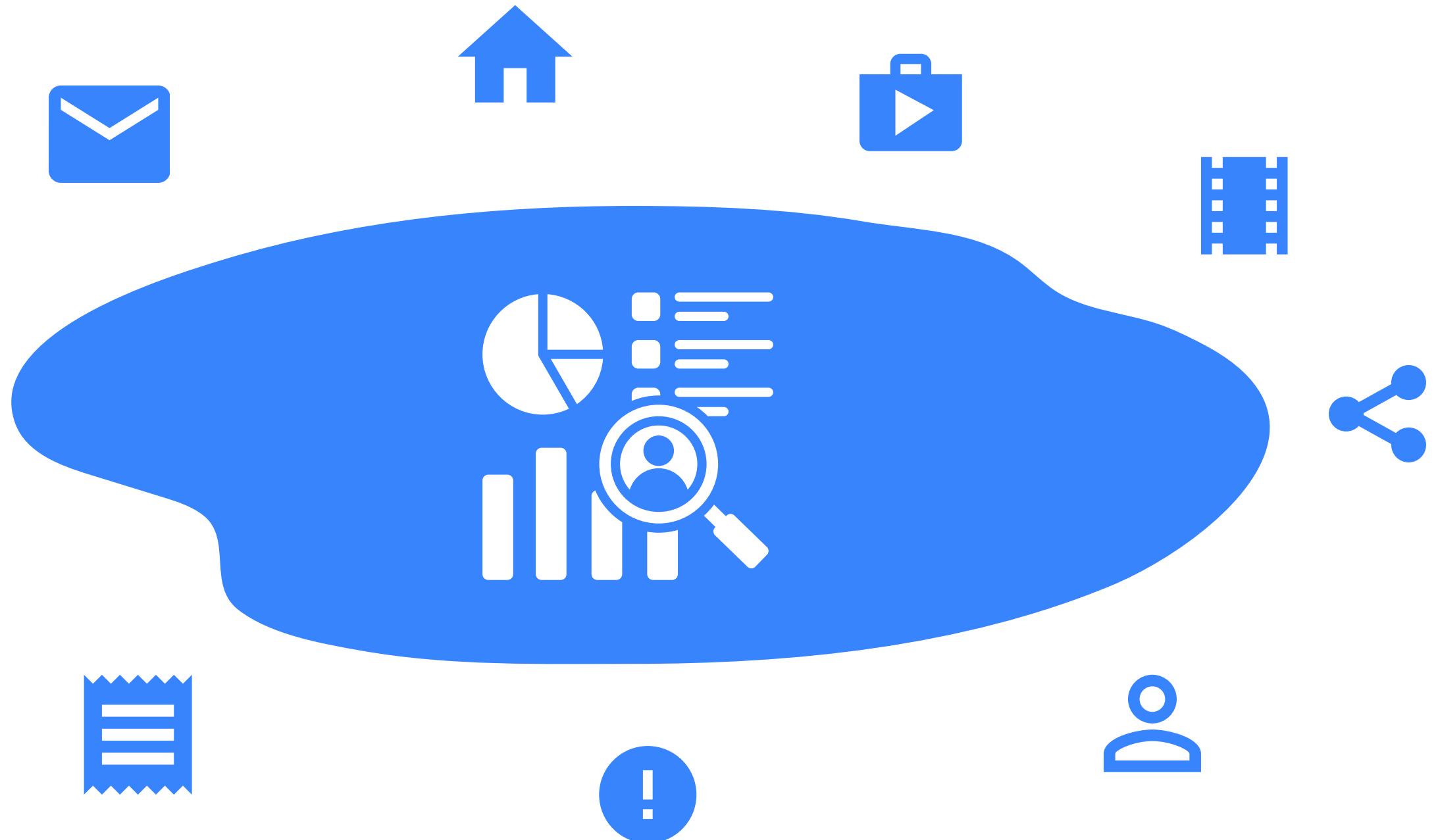
# Data Lake

Data Lake é um **conceito**, mas também pode ser considerado uma **arquitetura**.



# Data Lake

Serve como um **repositório central** para dados estruturados e não estruturados de diversas fontes para **análise** e tomada de **decisões**.



# Onde pode ser implementado?

- On-Premises
- Cloud
- Híbrido

# Quais seus componentes principais?

- Repositório de objetos
- Processamento de dados
- Bancos SQL e/ou NoSQL
- Análise e Visualização e etc...

# Onde pode ser implementado?

- On-Premises ←
- Cloud
- Híbrido

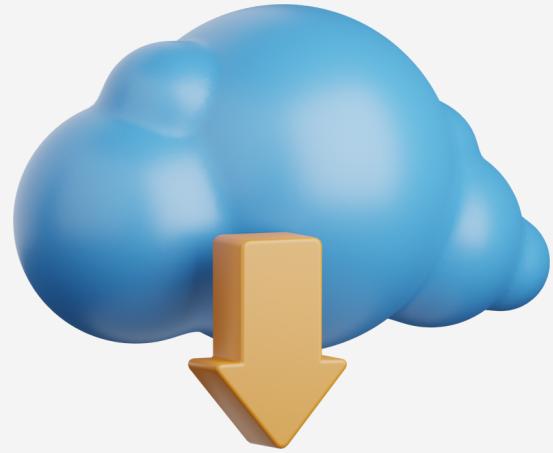


# Quais seus componentes principais?

- Repositório de objetos
- Processamento de dados
- Bancos SQL e/ou NoSQL
- Análise e Visualização e etc...

# Onde pode ser implementado?

- On-Premises
- Cloud
- Híbrido

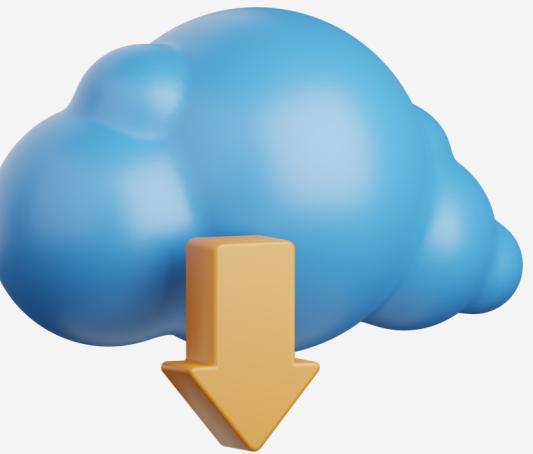


# Quais seus componentes principais?

- Repositório de objetos
- Processamento de dados
- Bancos SQL e/ou NoSQL
- Análise e Visualização e etc...

# Onde pode ser implementado?

- On-Premises
- Cloud
- Híbrido



# Quais seus componentes principais?

- Repositório de objetos
- Processamento de dados
- Bancos SQL e/ou NoSQL
- Análise e Visualização e etc...

# Onde pode ser implementado?

- On-Premises
- Cloud
- Híbrido



# Quais seus componentes principais?

- Repositório de objetos
- Processamento de dados
- Bancos SQL e/ou NoSQL
- Análise e Visualização e etc...

# Onde pode ser implementado?

- On-Premises
- Cloud
- Híbrido

# Quais seus componentes principais?

- Repositório de objetos
- Processamento de dados
- Bancos SQL e/ou NoSQL
- Análise e Visualização e etc...

# Onde pode ser implementado?

- On-Premises
- Cloud
- Híbrido

# Quais seus componentes principais?

- Repositório de objetos
- Processamento de dados
- Bancos SQL e/ou NoSQL
- Análise e Visualização e etc...



# Onde pode ser implementado?

- On-Premises
- Cloud
- Híbrido

# Quais seus componentes principais?

- Repositório de objetos
- Processamento de dados
- Bancos SQL e/ou NoSQL
- Análise e Visualização e etc...

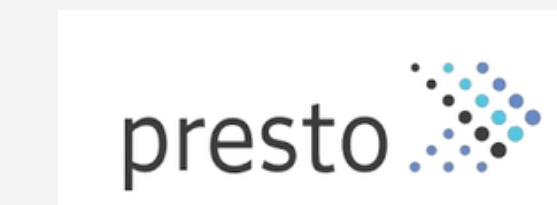


# Onde pode ser implementado?

- On-Premises
- Cloud
- Híbrido

# Quais seus componentes principais?

- Repositório de objetos
- Processamento de dados
- Bancos SQL e/ou NoSQL
- Análise e Visualização e etc...



# Onde pode ser implementado?

- On-Premises
- Cloud
- Híbrido

# Quais seus componentes principais?

- Repositório de objetos
- Processamento de dados
- Bancos SQL e/ou NoSQL
- Análise e Visualização e etc...



# Onde pode ser implementado?

- On-Premises
- Cloud
- Híbrido

# Quais seus componentes principais?

- Repositório de objetos
- Processamento de dados
- Bancos SQL e/ou NoSQL
- Análise e Visualização e etc...

# Onde pode ser implementado?

- On-Premises
- Cloud
- Híbrido

# Quais seus componentes principais?

- Repositório de objetos
- Processamento de dados
- Bancos SQL e/ou NoSQL
- Análise e Visualização e etc...



Fonte: <https://theinfinityconferences.wordpress.com/2017/10/05/top-10-open-source-big-data-tools/>

# Data Lake

Ele pode ser organizado em **camadas**.



# Data Lake

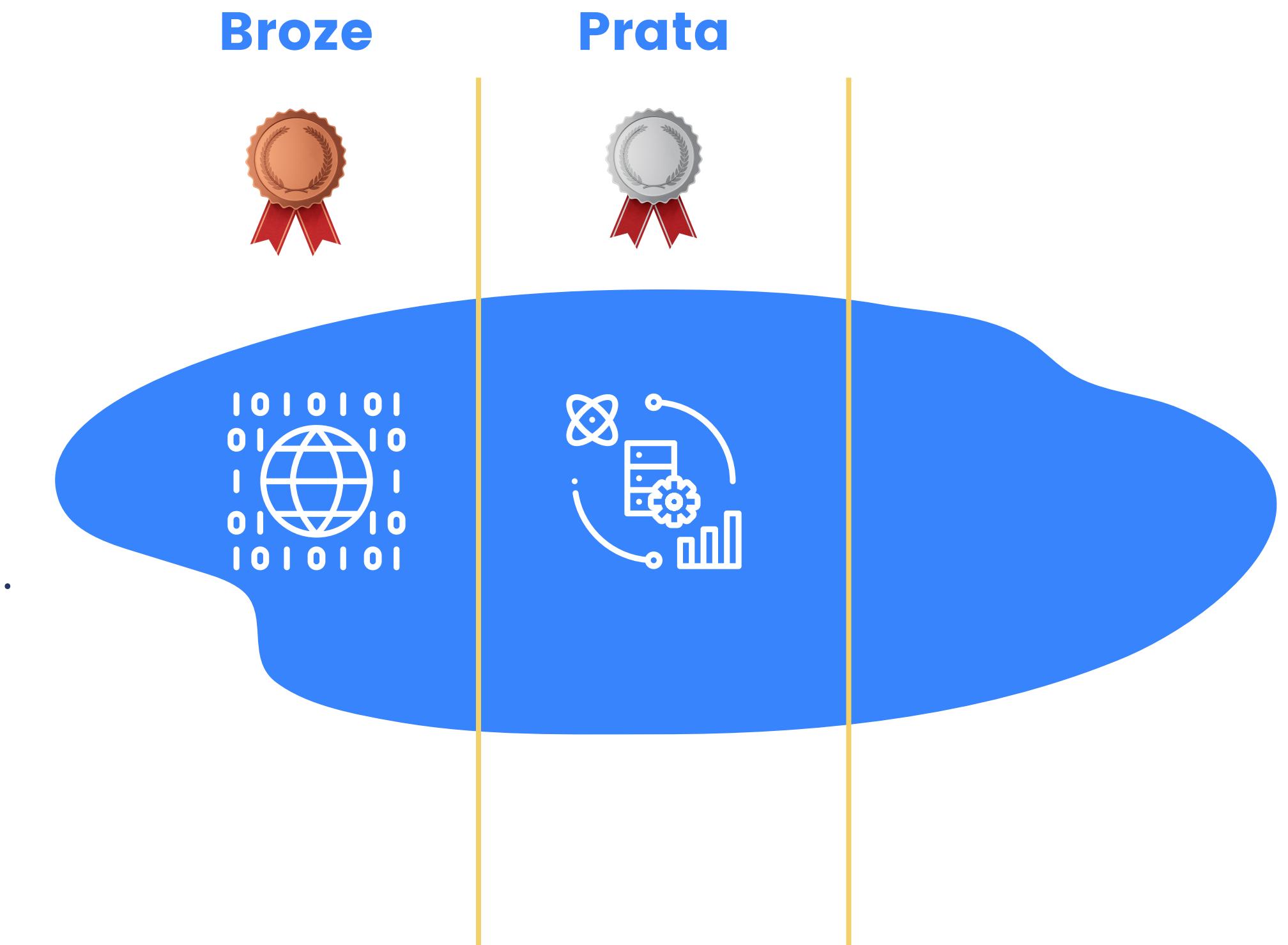
Ele pode ser organizado em **camadas**.

**Broze**



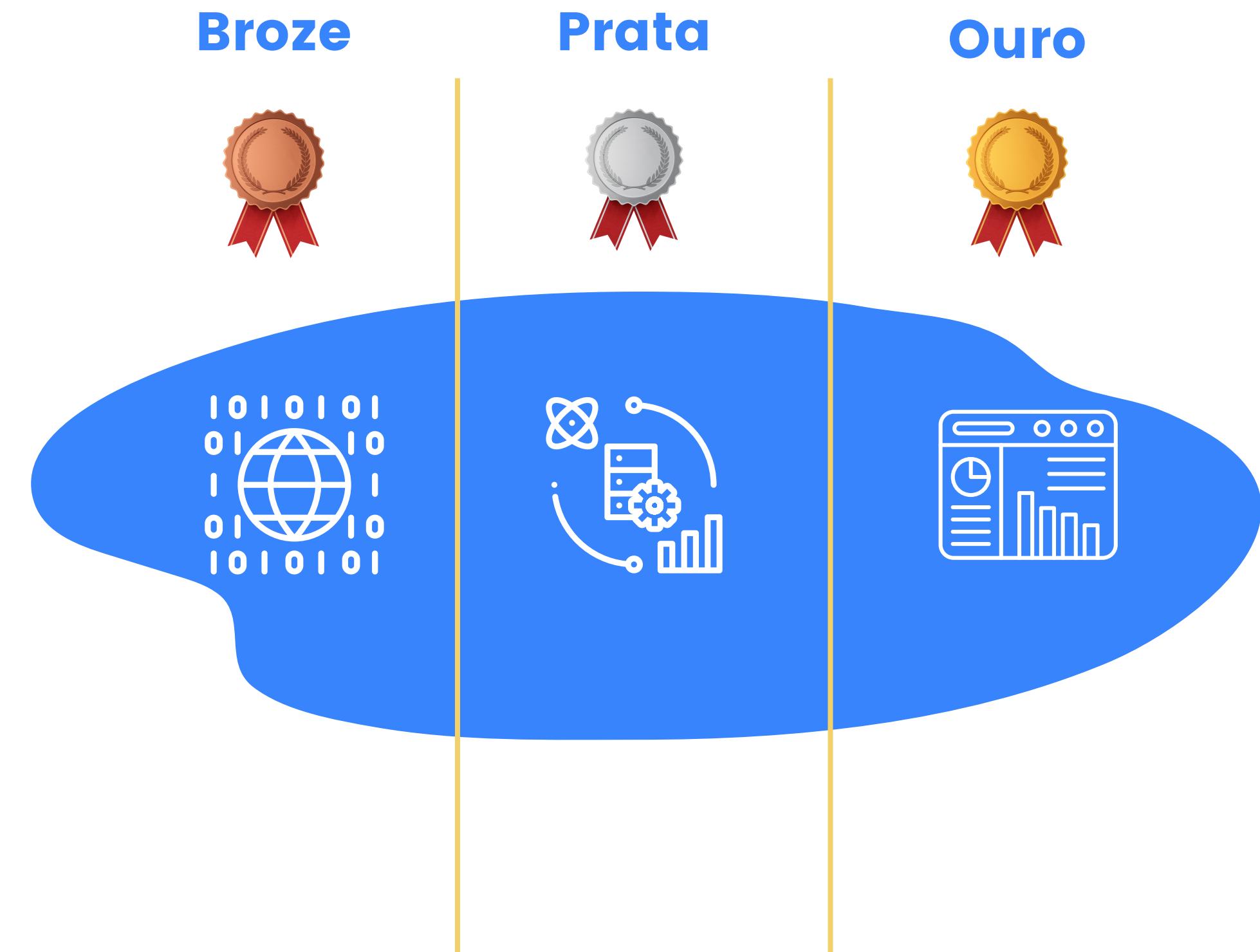
# Data Lake

Ele pode ser organizado em **camadas**.



# Data Lake

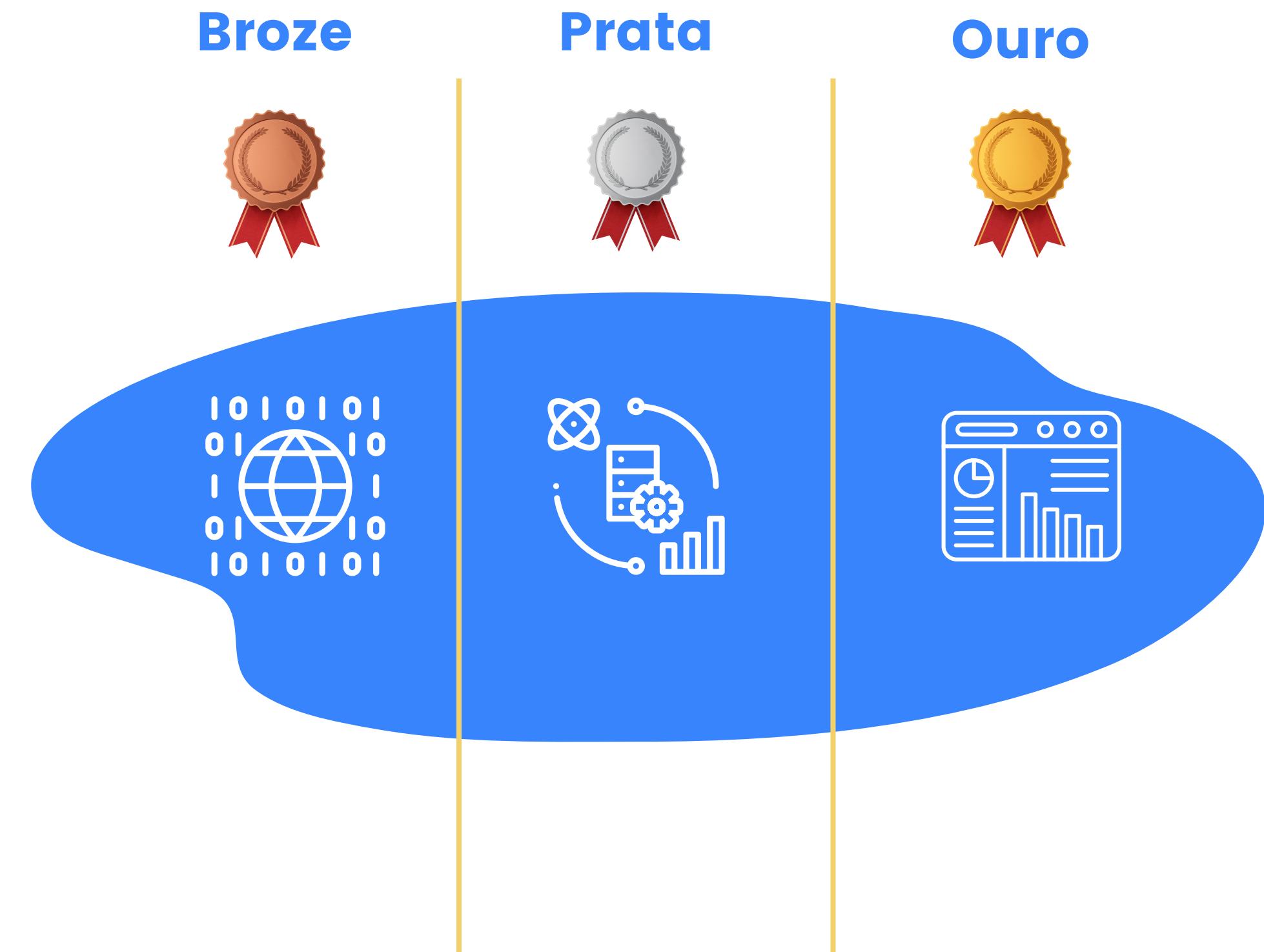
Ele pode ser organizado em **camadas**.



## Medallion Architecture

### Data Lake

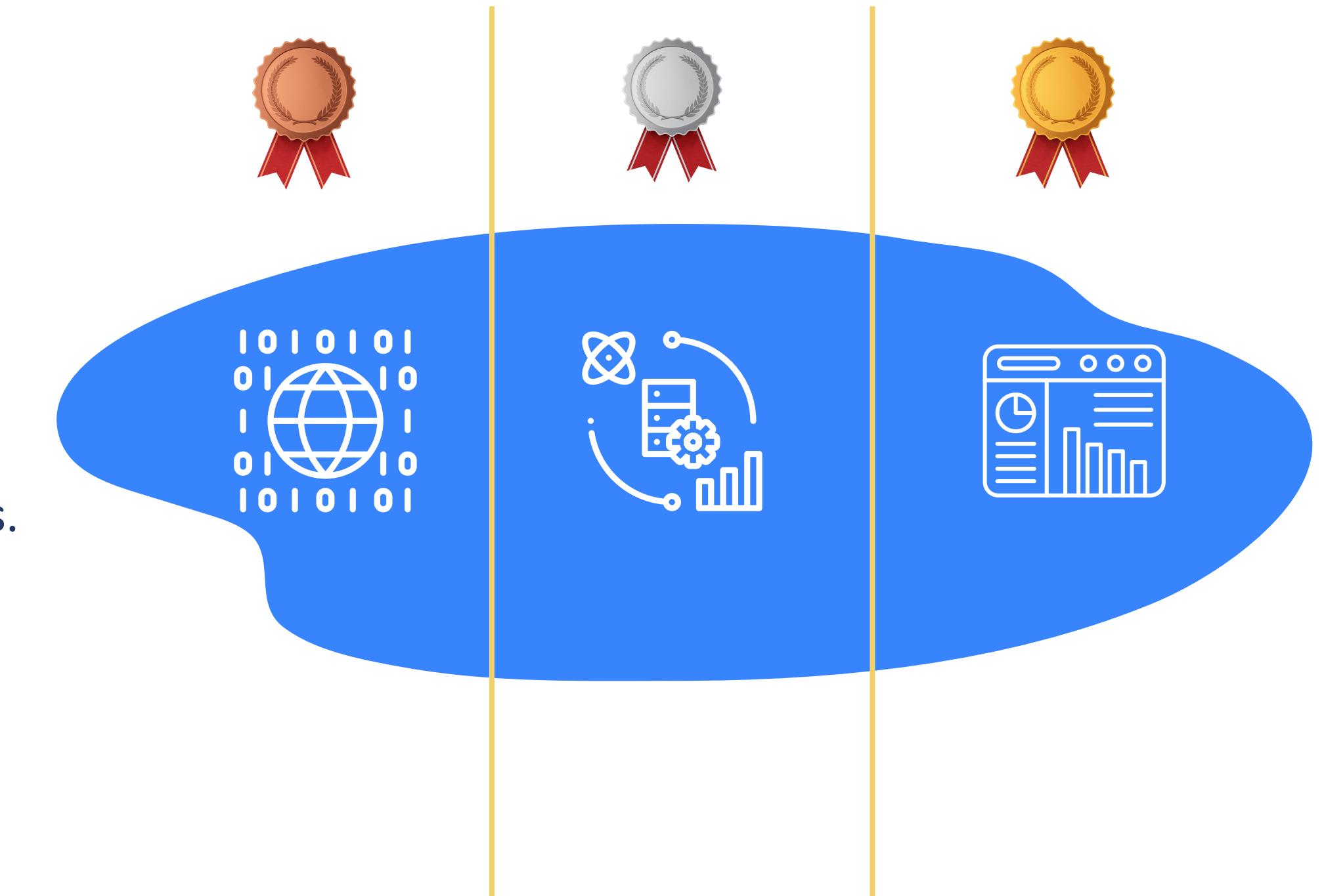
Ele pode ser organizado em **camadas**.



## Lógica ou Física!?

### Data Lake

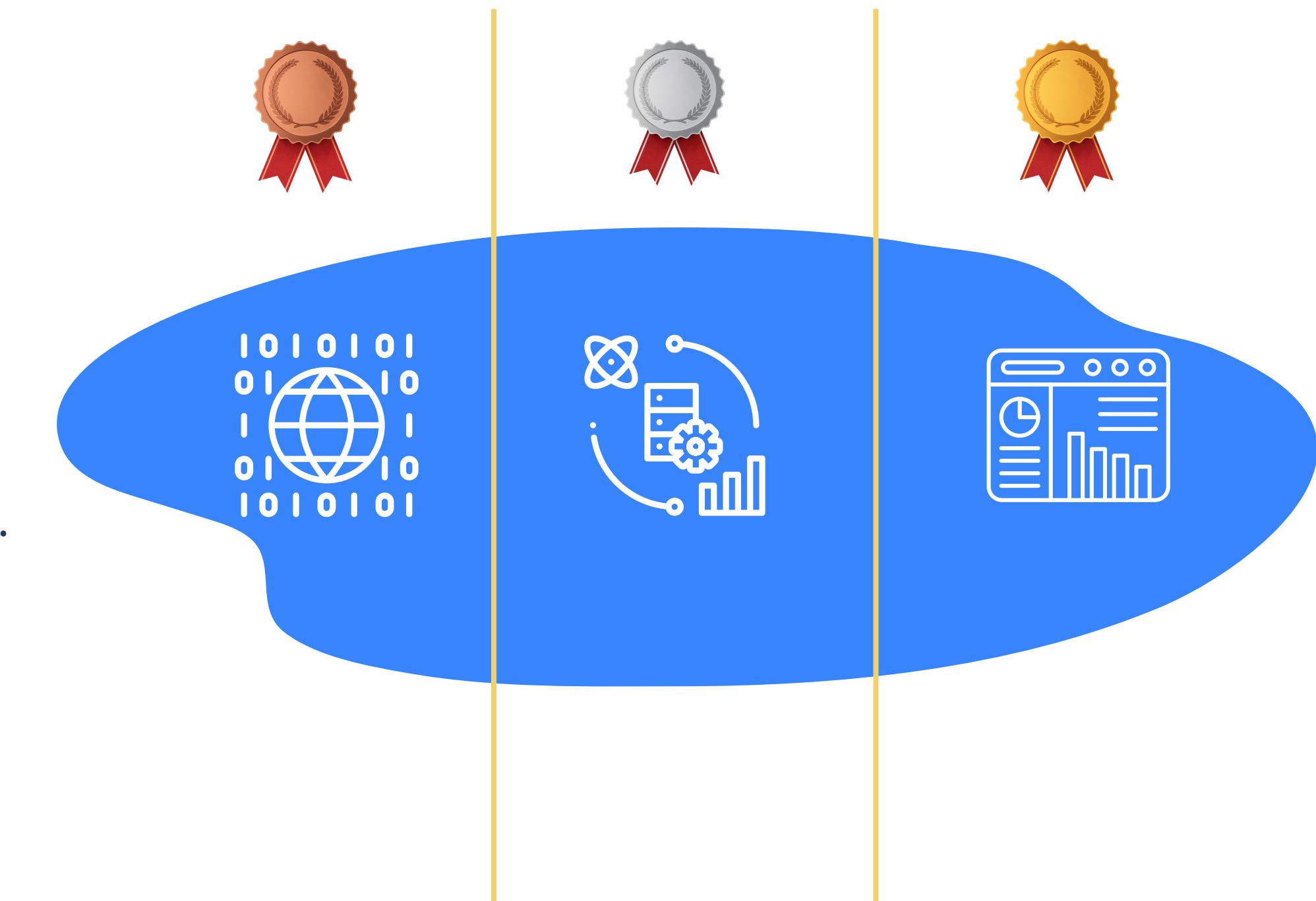
Ele pode ser **organizado** em camadas.



# Data Lake

Ele pode ser organizado em camadas.

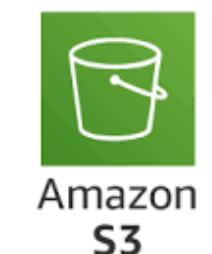
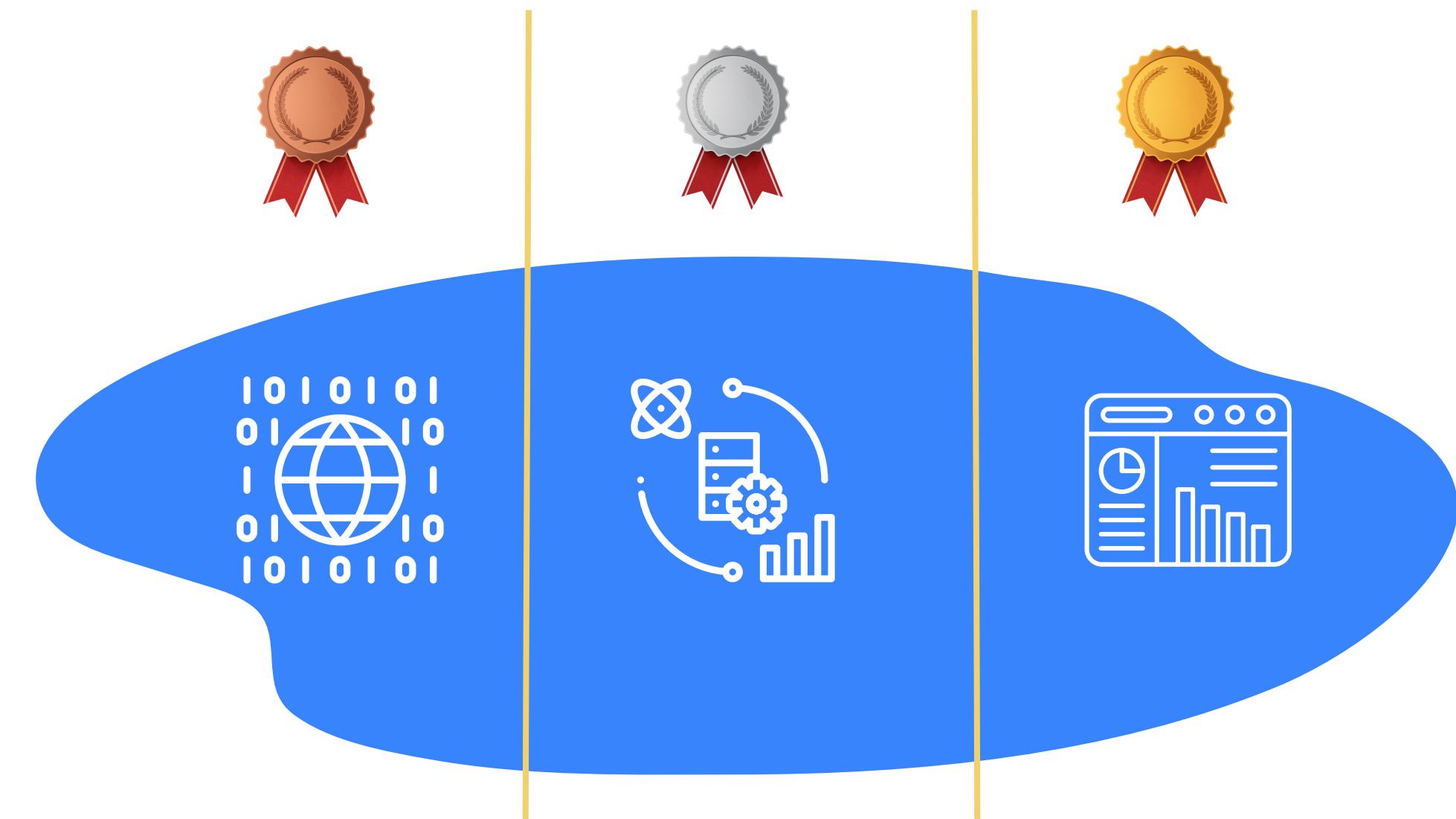
Lógica



# Data Lake

Ele pode ser organizado em camadas.

## Lógica



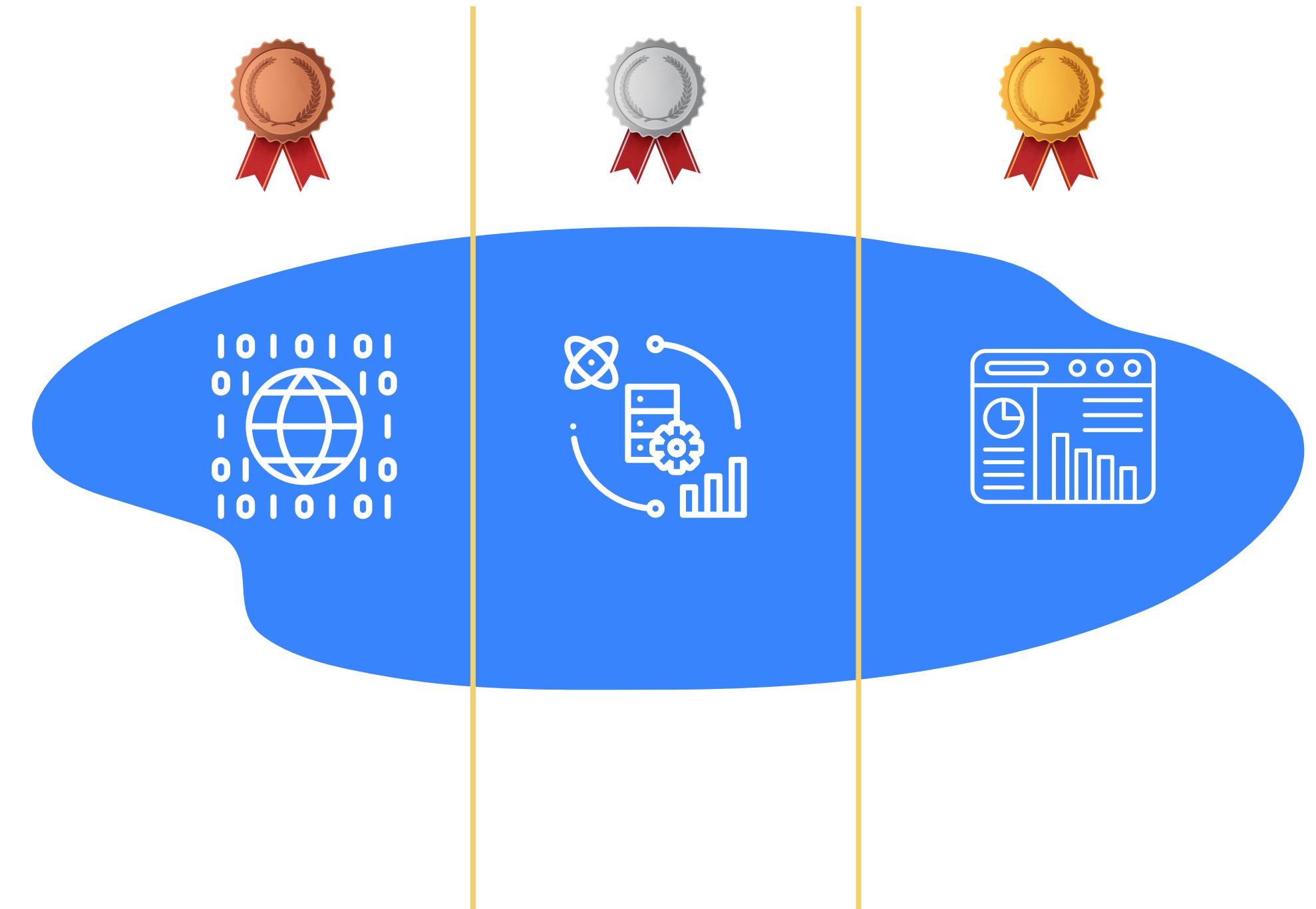
Amazon  
S3



# Data Lake

Ele pode ser organizado em camadas.

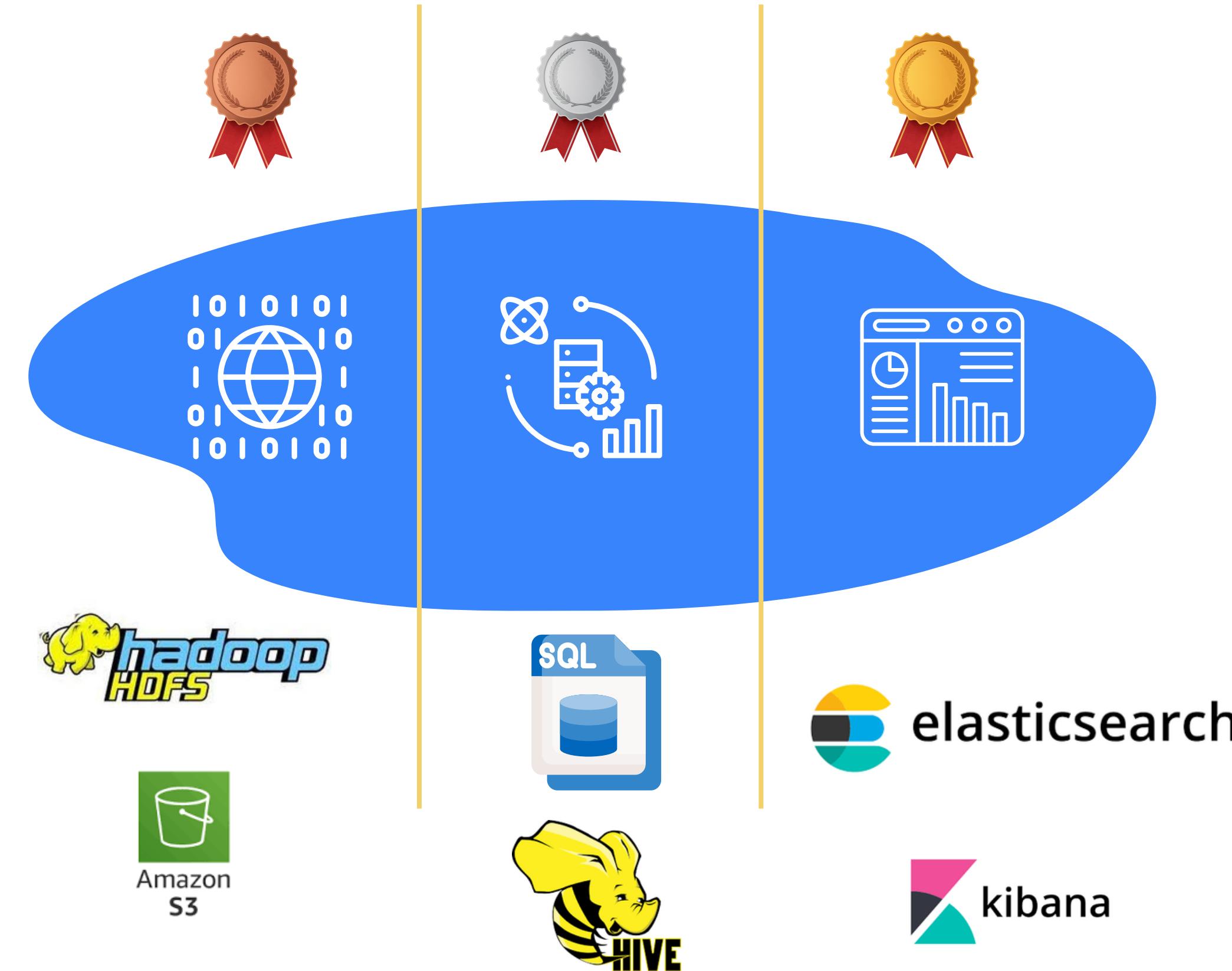
**Física**



# Data Lake

Ele pode ser organizado em camadas.

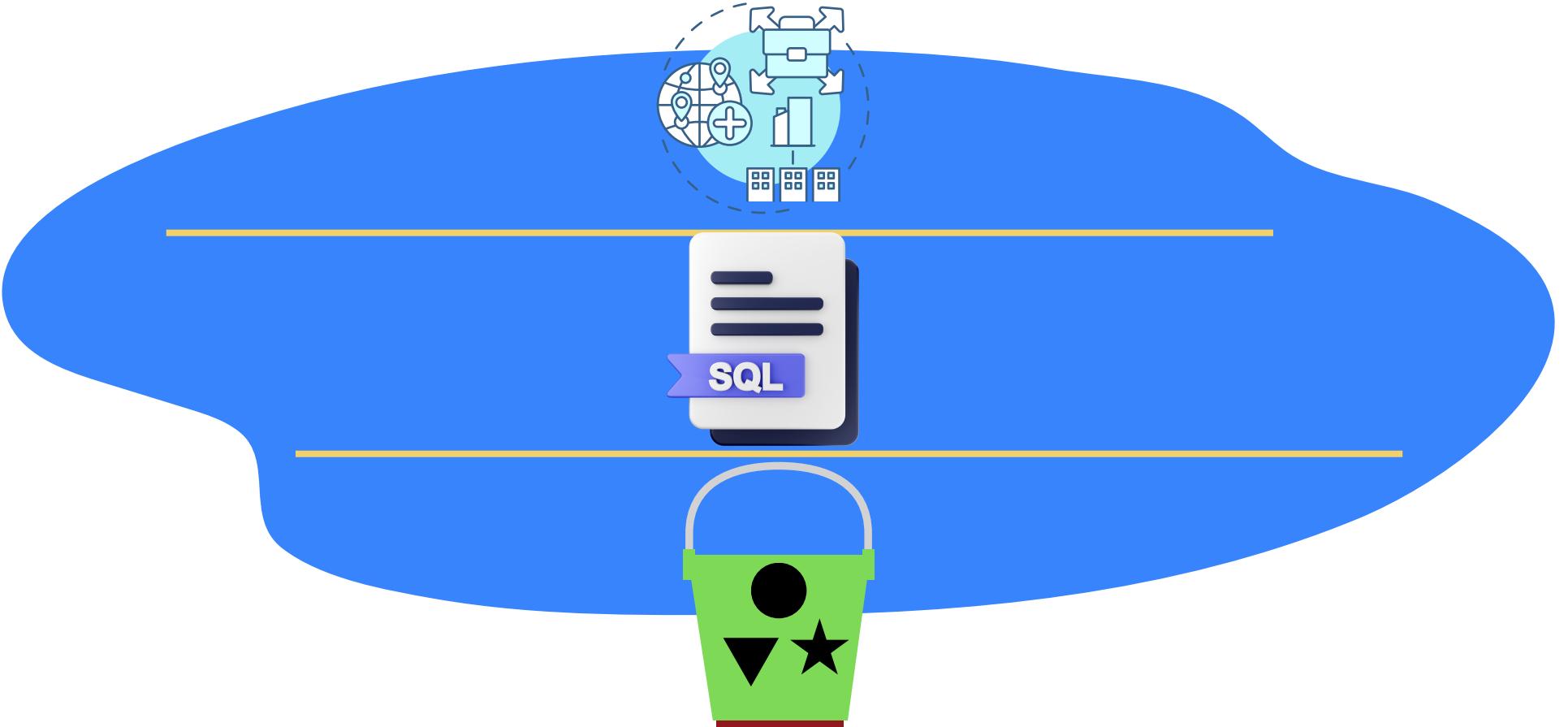
Física



# Data Lake

Na prática...

**Camadas!**



# Data Lake

Na prática...

**Camadas!**

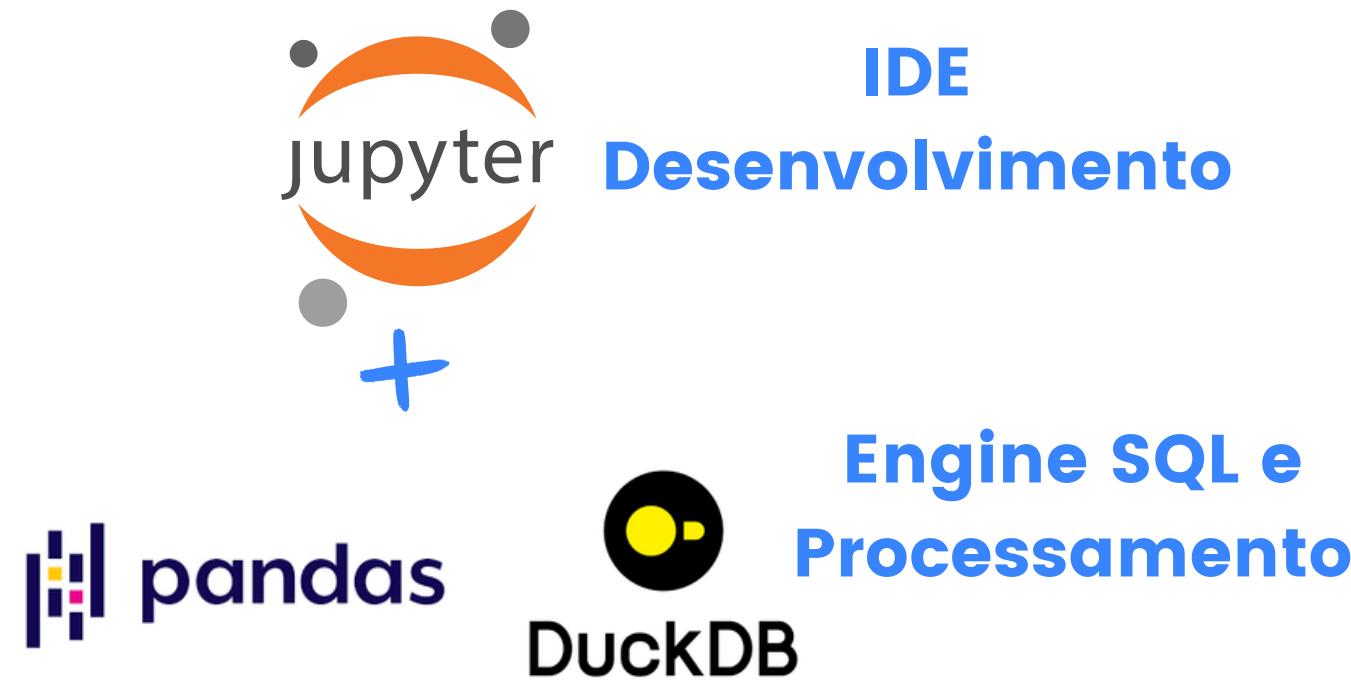


**IDE**  
**Desenvolvimento**



# Data Lake

Na prática...



**Camadas!**



# Data Lake

Na prática...



**Camadas!**



# Data Lake

Na prática...



**Camadas!**

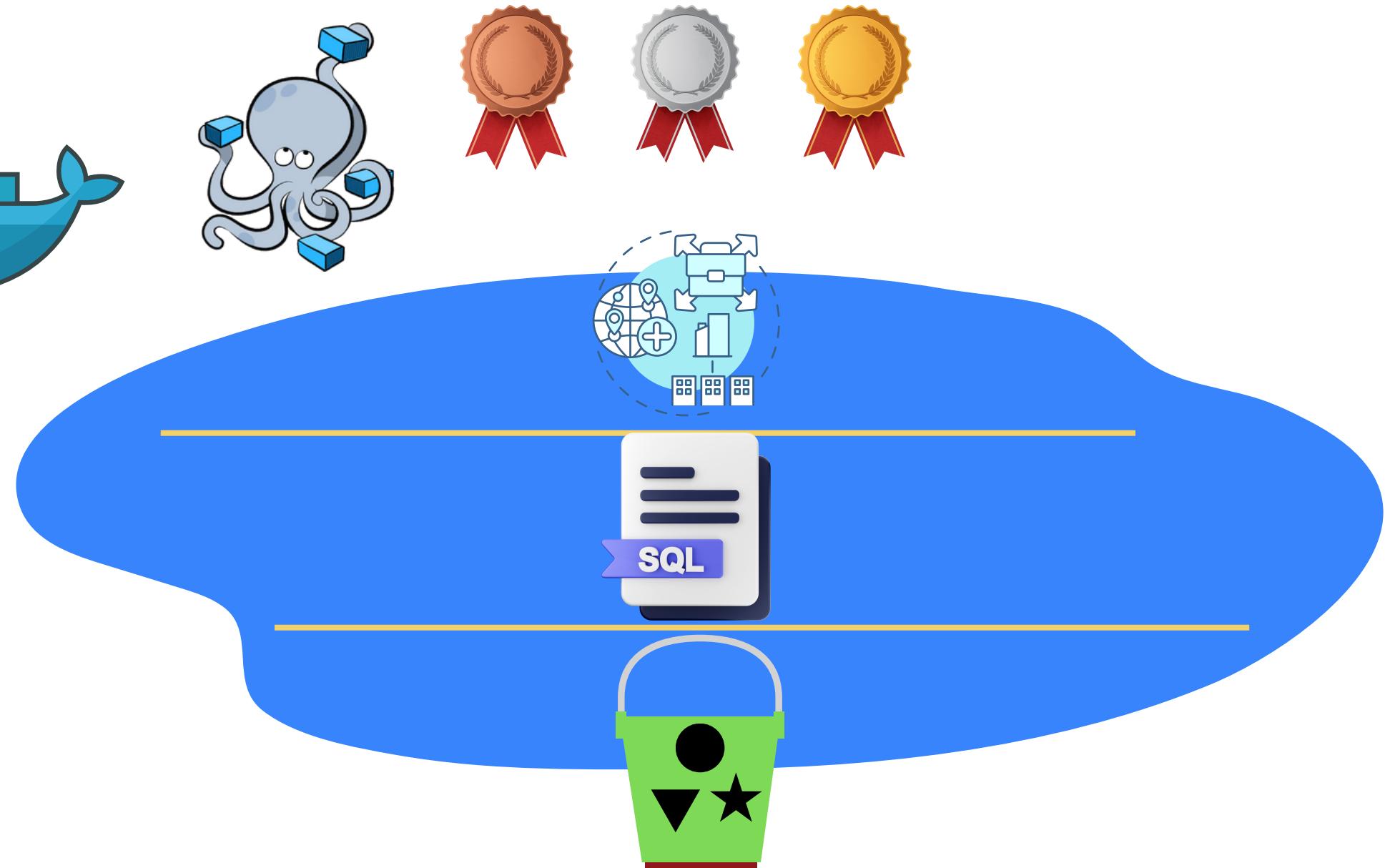


# Data Lake

Na prática...



**Camadas!**

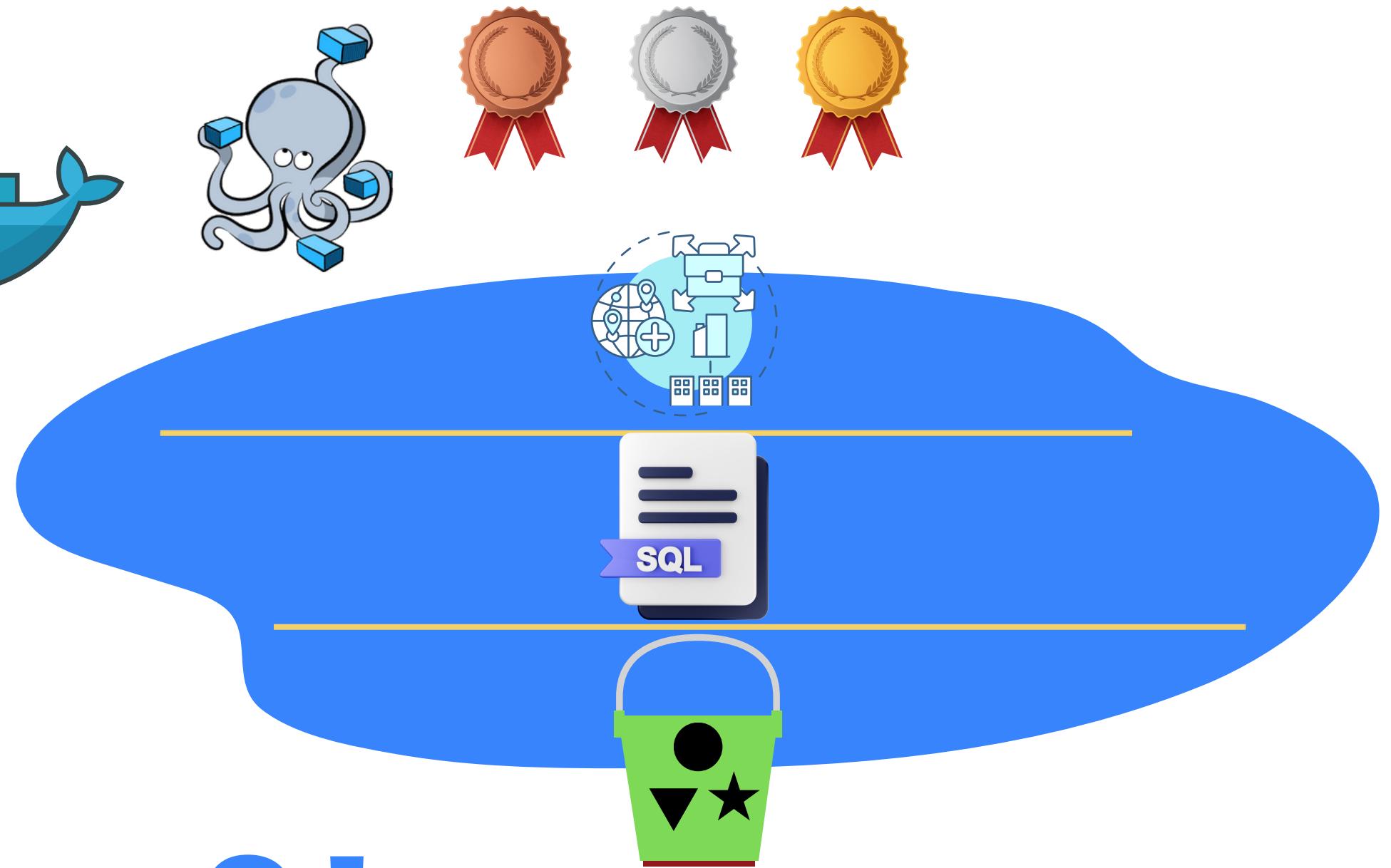


# Data Lake

Na prática...



**Camadas!**



**Vamos?!**

# Referências:



Data Lake, **Wikipédia**. Disponível em:  
[https://pt.wikipedia.org/wiki/Data\\_Lake](https://pt.wikipedia.org/wiki/Data_Lake). Acesso em: 22 out. de 2023.

DATA LAKE: SAIBA COMO FUNCIONA A TECNOLOGIA DE ARMAZENAMENTO DE DADOS, **Tera Blog**. Disponível em:  
<https://blog.somostera.com/data-science/data-lake>. Acesso em: 25 out. de 2023.

Unlock the full potential of a modern data warehouse with Delta Lake, **Blueprint**. Disponível em: <https://bpcs.com/blog/why-delta-lake>. Acesso em: 25 out. de 2023.

Medallion Architecture, **Data Engineering**. Disponível em:  
<https://dataengineering.wiki/Concepts/Medallion+Architecture>. Acesso em: 25 out. de 2023.

O que é um data lake?, **AWS**. Disponível em:  
<https://aws.amazon.com/pt/big-data/datalakes-and-analytics/what-is-a-data-lake/>. Acesso em: 28 out. de 2023.