

Crie seu pipeline de dados na cloud sem pagar nada por isso!

Erika Nagamine

Trilha Inovação com dados em nuvem

Crie seu pipeline de dados na cloud sem pagar nada por isso!

Erika Nagamine

Trilha Inovação com dados em nuvem



Este trabalho está licenciado sob uma Licença Creative Commons Atribuição-Compartilhável 4.0 Internacional. Para ver uma cópia desta licença, visite <http://creativecommons.org/licenses/by-sa/4.0/>.

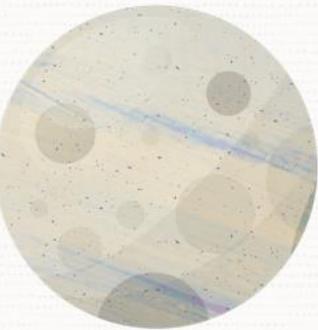
\$> whoami



Erika Nagamine

Principal Enterprise Cloud Solutions Architect

Data Management | Data Engineer | Data Scientist | Analytics



<https://linkedin.com/in/erikanagamine>



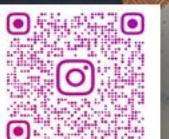
<https://github.com/erikanagamine>



@erikanagamine



@erikanagamine

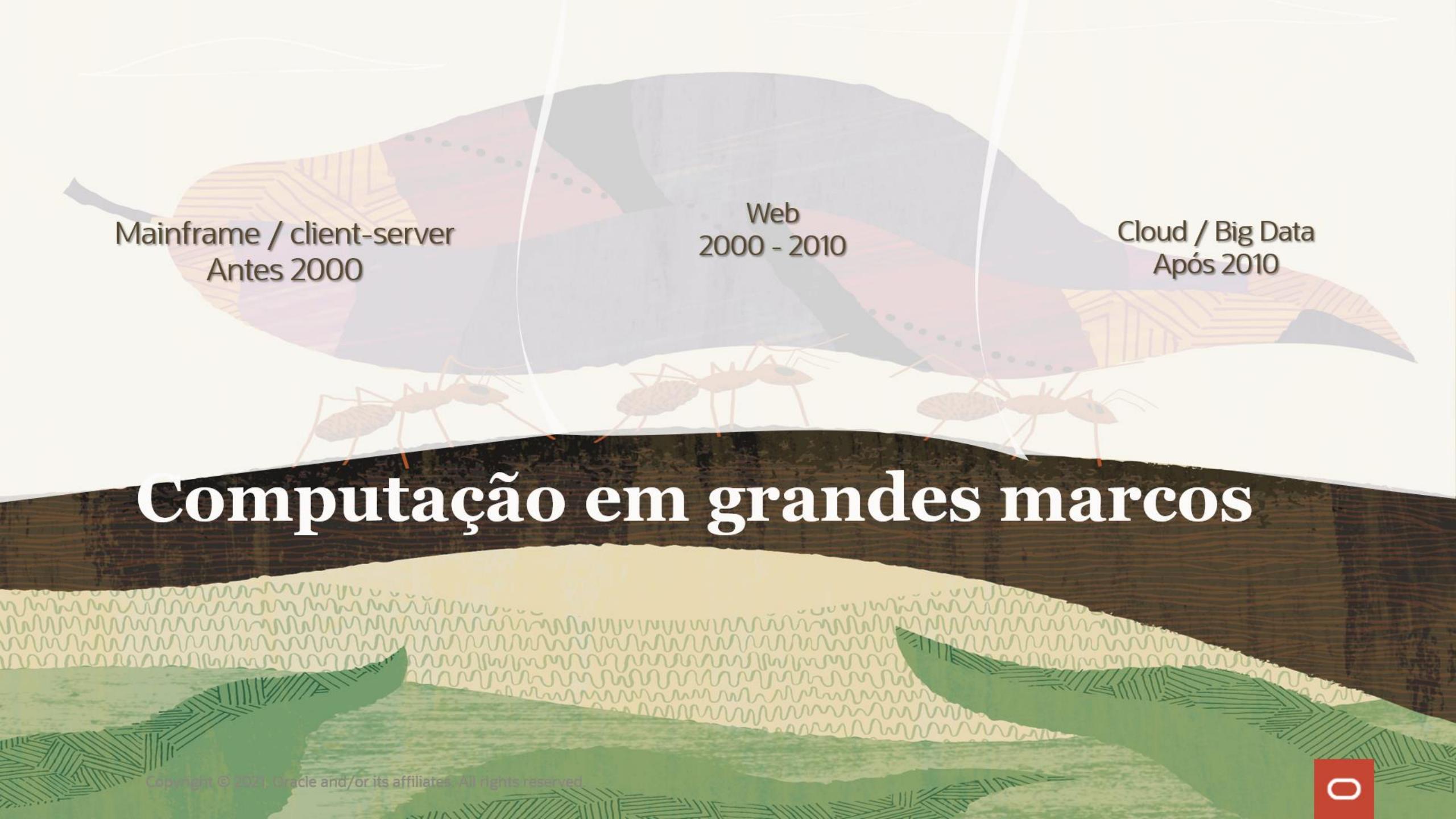


ERIKANAGAMINE



Dados estão mudando o mundo



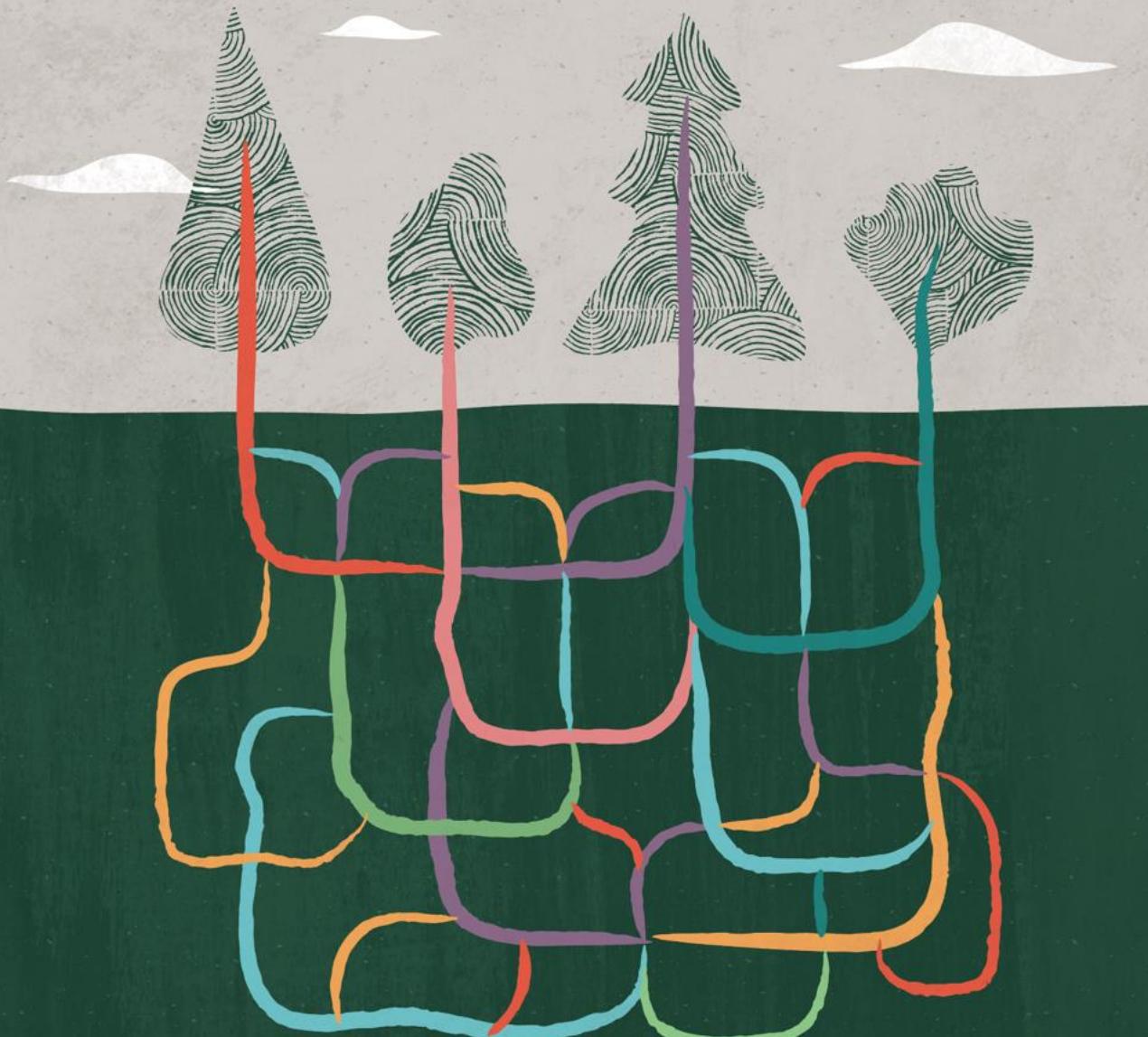


Mainframe / client-server
Antes 2000

Web
2000 - 2010

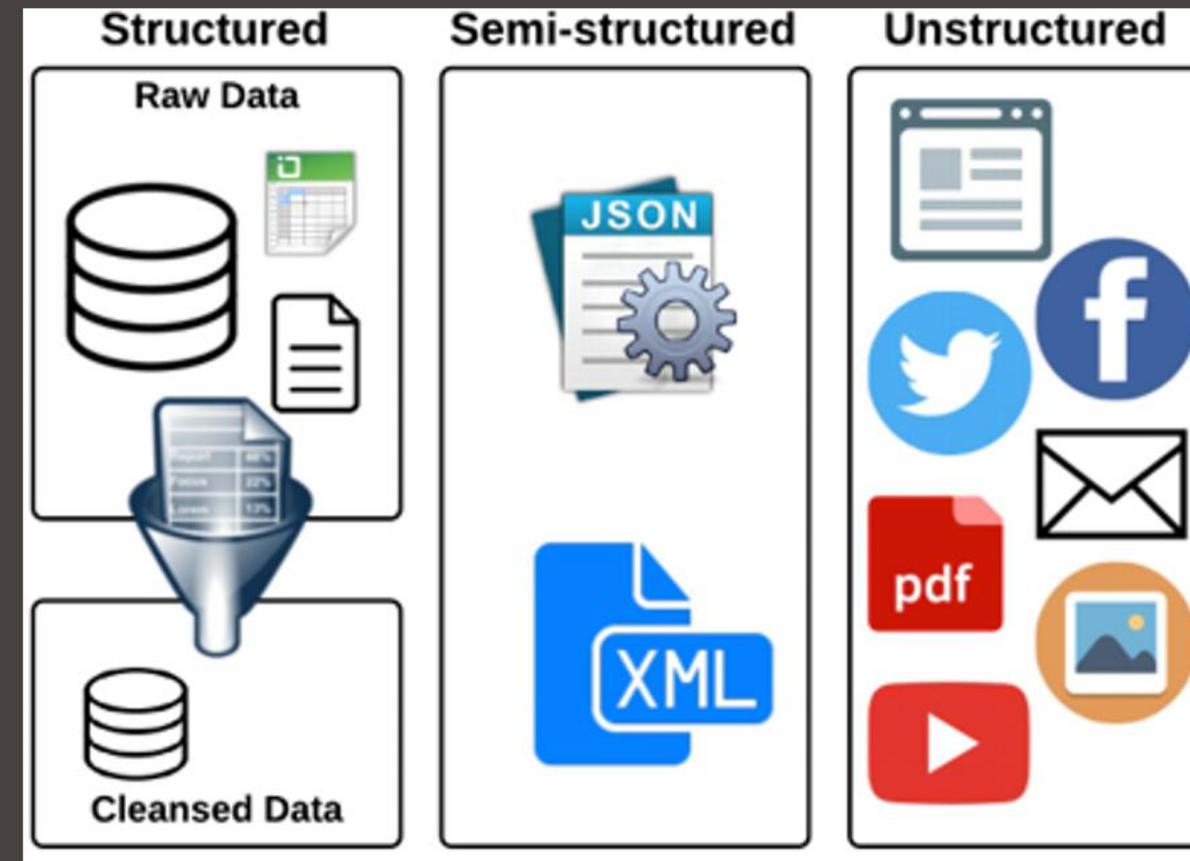
Cloud / Big Data
Após 2010

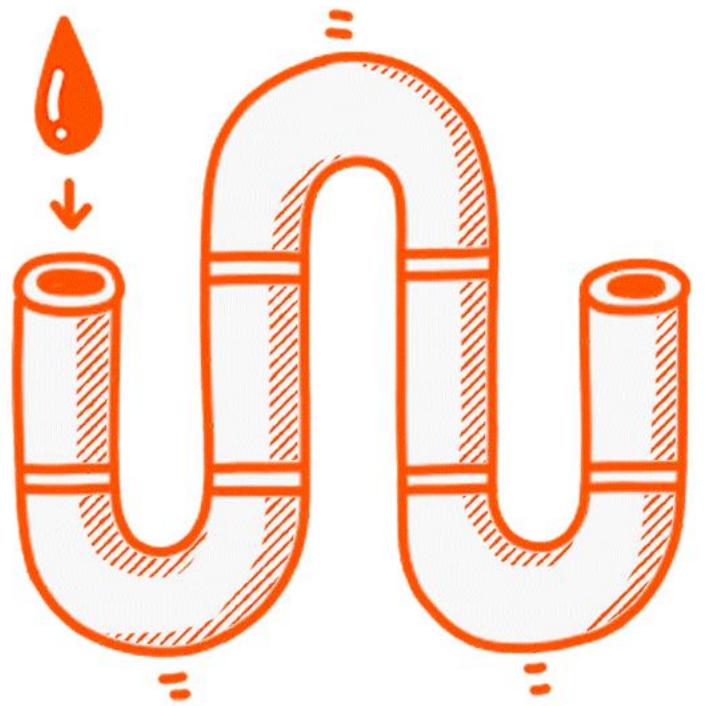
Computação em grandes marcos



Num mundo onde
cada vez geramos
mais dados, analisar
e correlacionar
pode ser mais
complexo

Tipos de dados



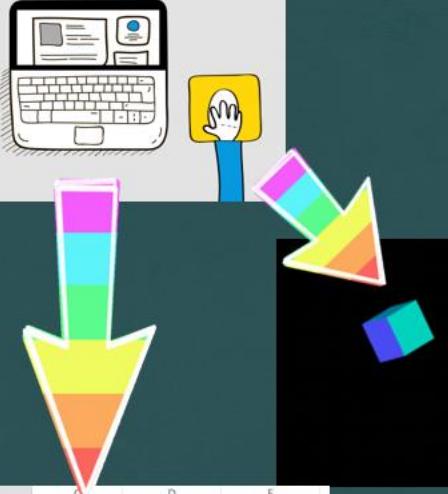


Data Pipeline

Na computação, um pipeline, também conhecido como **pipeline de dados**, é um conjunto de elementos de processamento de dados conectados em série, onde a saída de um elemento é a entrada do próximo. Os elementos de um pipeline são geralmente executados em paralelo ou em batch.

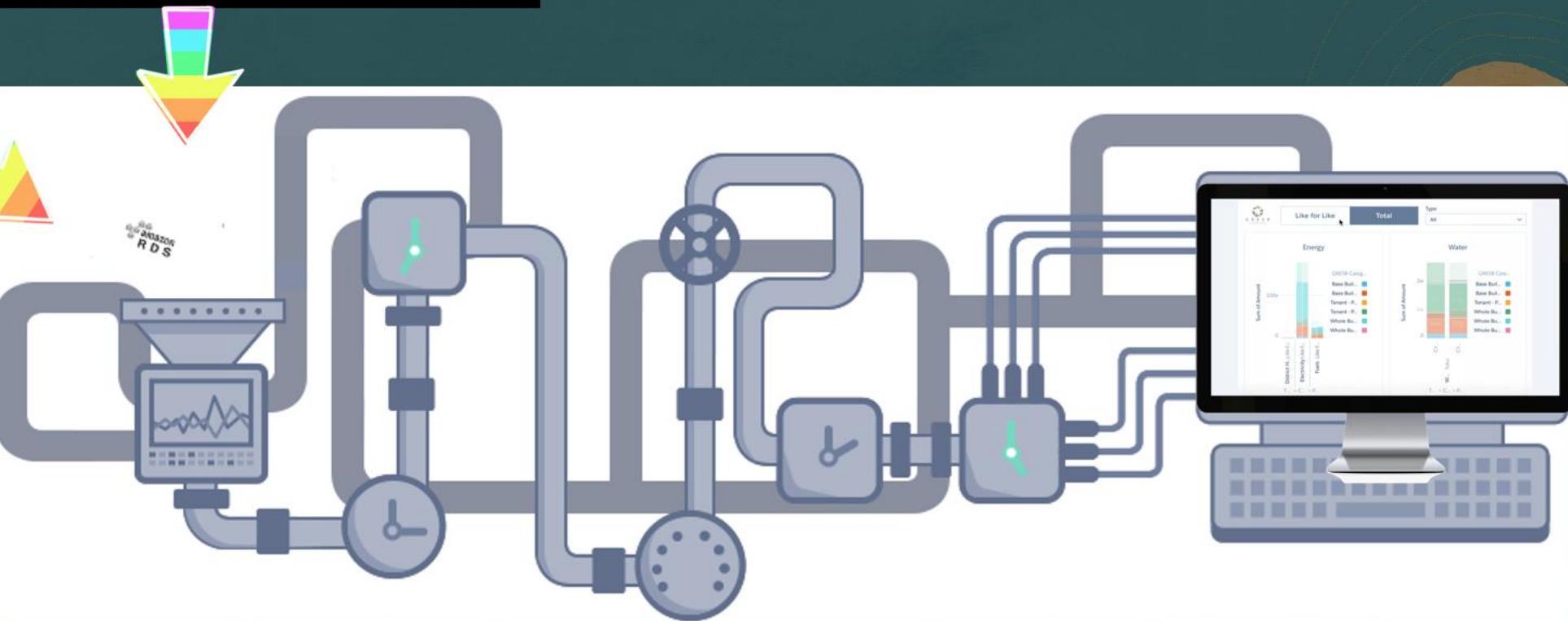
- Wikipedia





SDK Software Development Kit

A	B	C	D	E	
1	Vendedor	Janeiro	Fevereiro	Março	Abril
2	Marcio	R\$ 1.300,00	R\$ 2.890,00	R\$ 3.500,00	R\$ 5.400,00
3	Bruno	R\$ 1.670,00	R\$ 2.900,00	R\$ 5.435,00	R\$ 4.676,00
4	Guilherme	R\$ 3.890,00	R\$ 3.900,00	R\$ 3.253,00	R\$ 5.446,00
5	Jonas	R\$ 3.409,00	R\$ 4.503,00	R\$ 2.554,00	R\$ 4.435,00
6	Bianca	R\$ 3.400,00	R\$ 3.563,00	R\$ 3.366,00	R\$ 6.655,00
7	Andreia	R\$ 2.900,00	R\$ 5.335,00	R\$ 3.246,00	R\$ 3.356,00
8					
9					
10					
11	Vendedor	Jonas			
12	Mês	Marco			
13	Valor	=DESC			
14			Retorna a taxa de desconto de um título		
15					
16					
17					



Fonte: <https://gfycat.com/foolishprestigioussharpseal>

<https://www.google.com/url?sa=i&url=https%3A%2F%2Fbddlabs.com%2Fhome%2F&psig=AOvVaw1DjfYuMk16VuizJQ0AzOjt&ust=1616509756085000&source=images&cd=vfe&ved=0CA0QjhxqFwoTCKC6yKOOxO8CFQAAAAAdAAAAABAJ>

Big Data

Na era da geração massiva de dados, o que é?

Big Data é um termo para *conjunto de dados que são tão grandes ou complexos que aplicações de processamento de dados tradicionais ficam inadequadas para tratar com eles.* O termo “big data” constantemente se refere a simplicidade de usar análises preditivas, análise de sentimentos de usuários, ou em certos momentos outros métodos analíticos para extrair valor dos dados e raramente de um conjunto específico de dados.

- Wikipedia

Big Data

Value

Visualization

Variety

Velocity

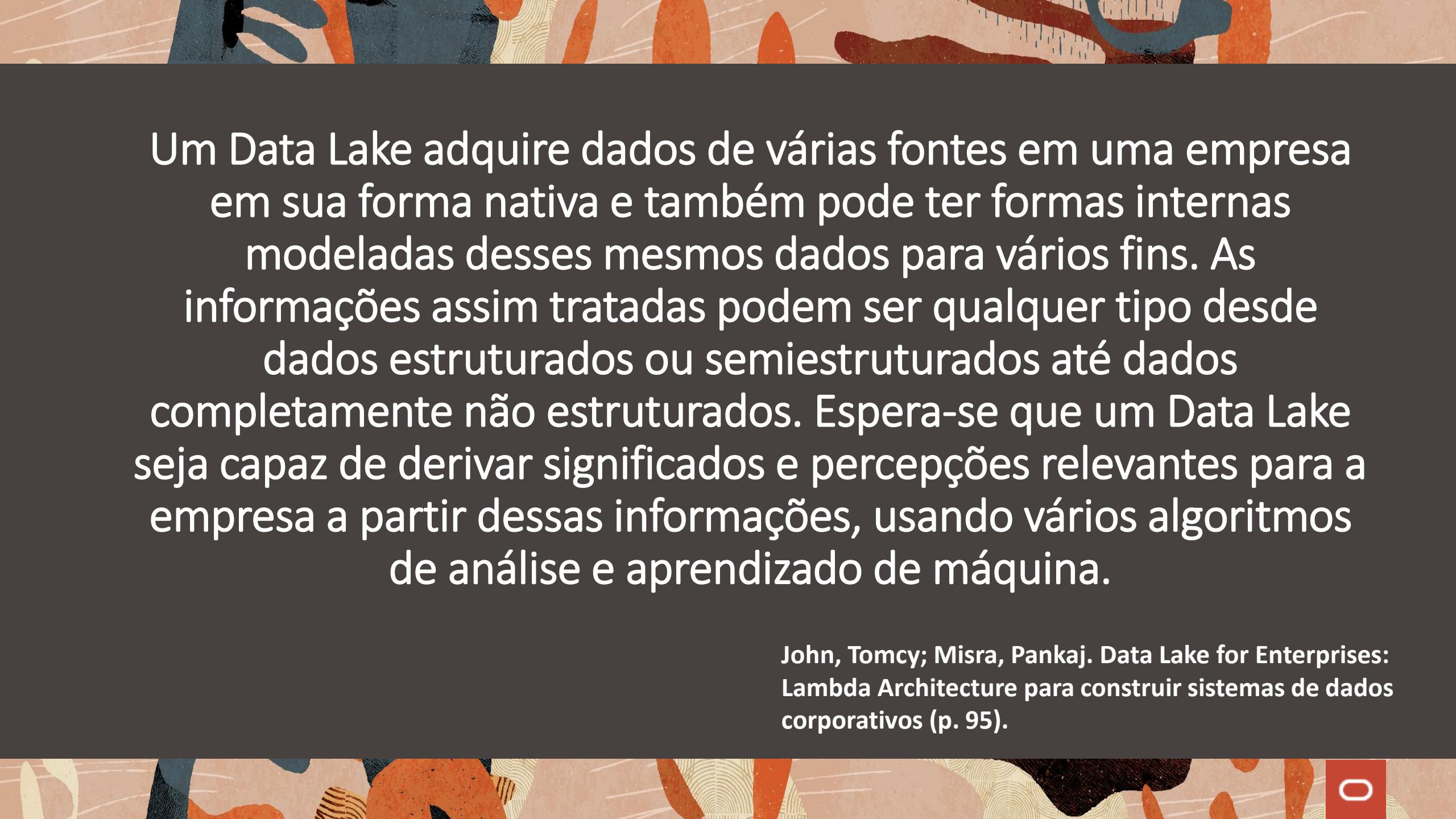
Volume

Veracity

Vision

Data Lake

Armazenar e extrair valor do seu dado pode ser uma tarefa árdua...



Um Data Lake adquire dados de várias fontes em uma empresa em sua forma nativa e também pode ter formas internas modeladas desses mesmos dados para vários fins. As informações assim tratadas podem ser qualquer tipo desde dados estruturados ou semiestruturados até dados completamente não estruturados. Espera-se que um Data Lake seja capaz de derivar significados e percepções relevantes para a empresa a partir dessas informações, usando vários algoritmos de análise e aprendizado de máquina.

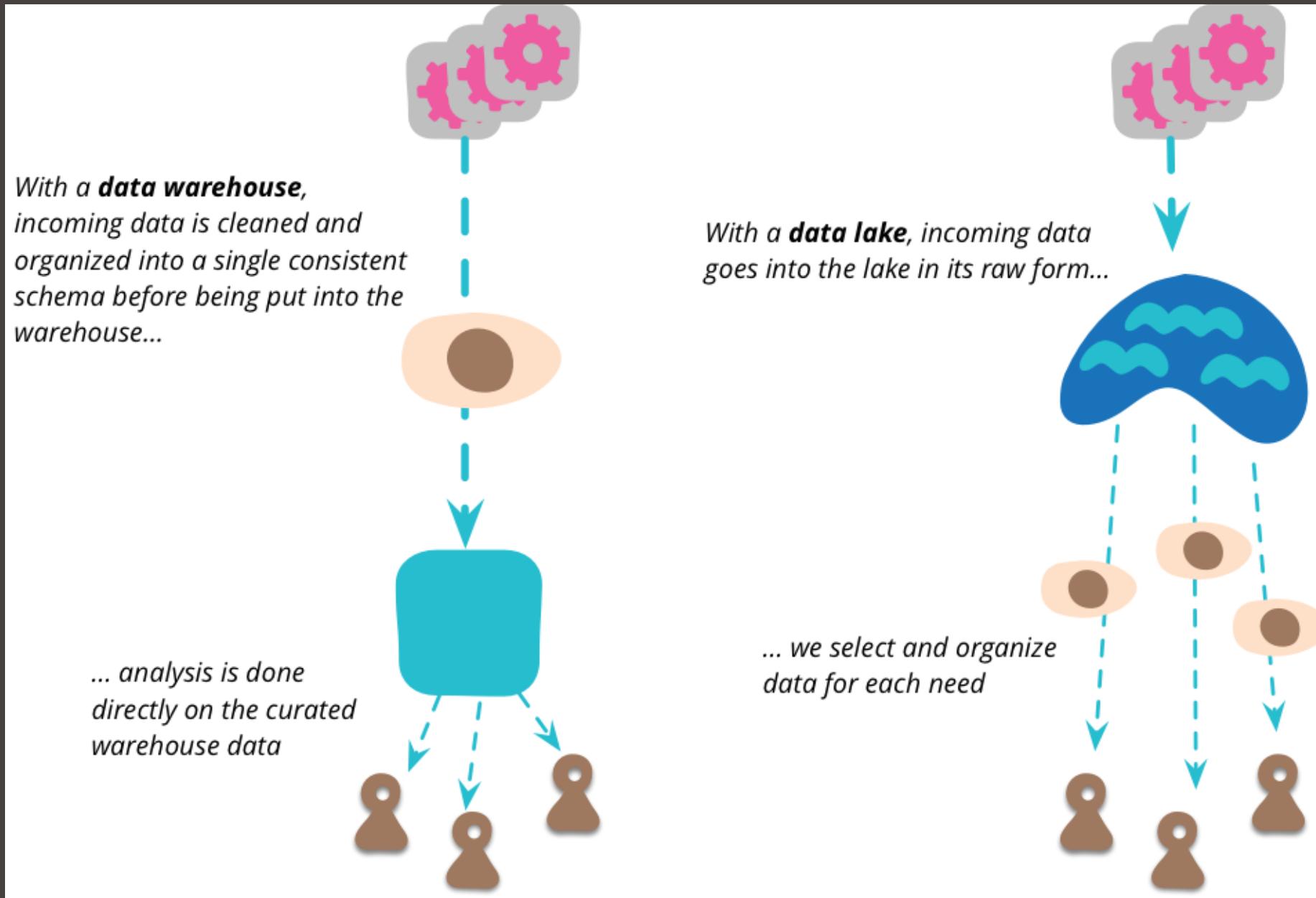
John, Tomcy; Misra, Pankaj. *Data Lake for Enterprises: Lambda Architecture para construir sistemas de dados corporativos* (p. 95).

Dado é ativo

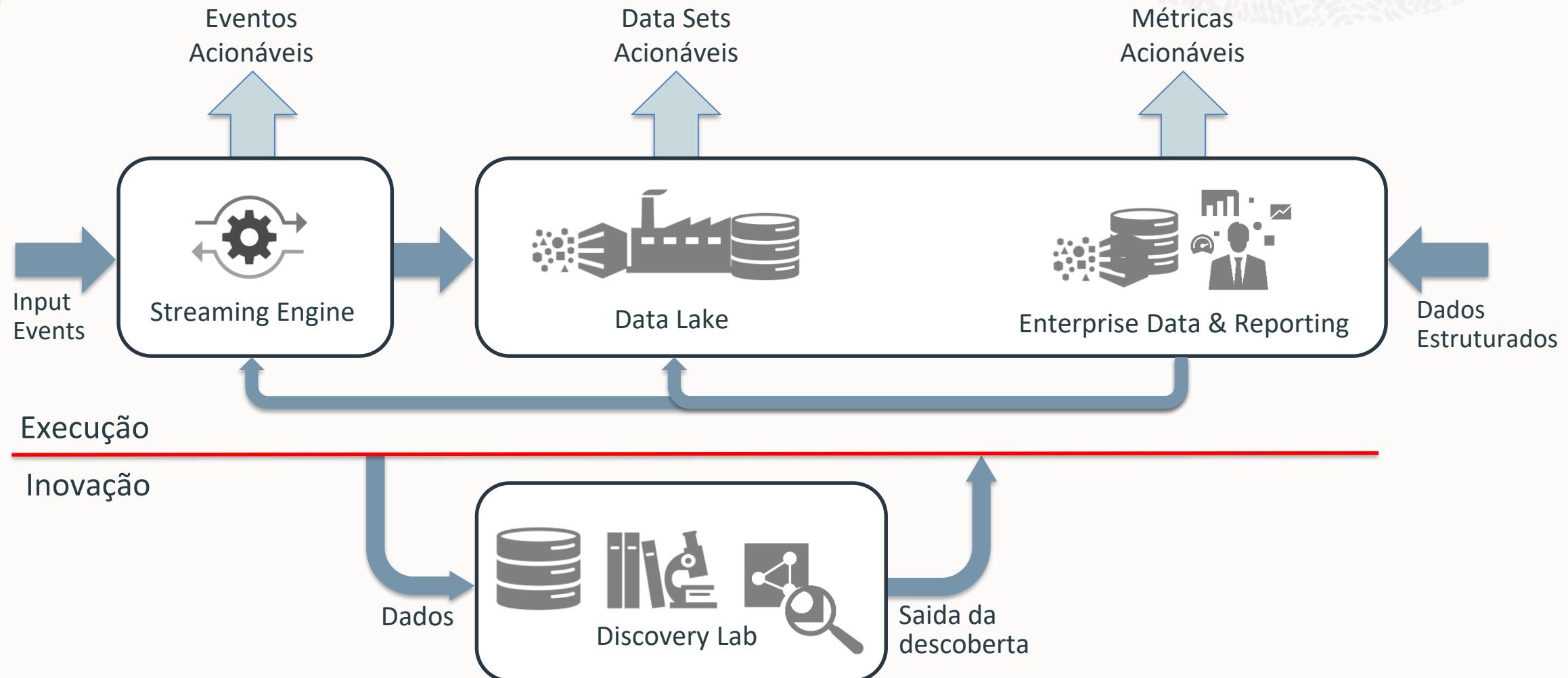




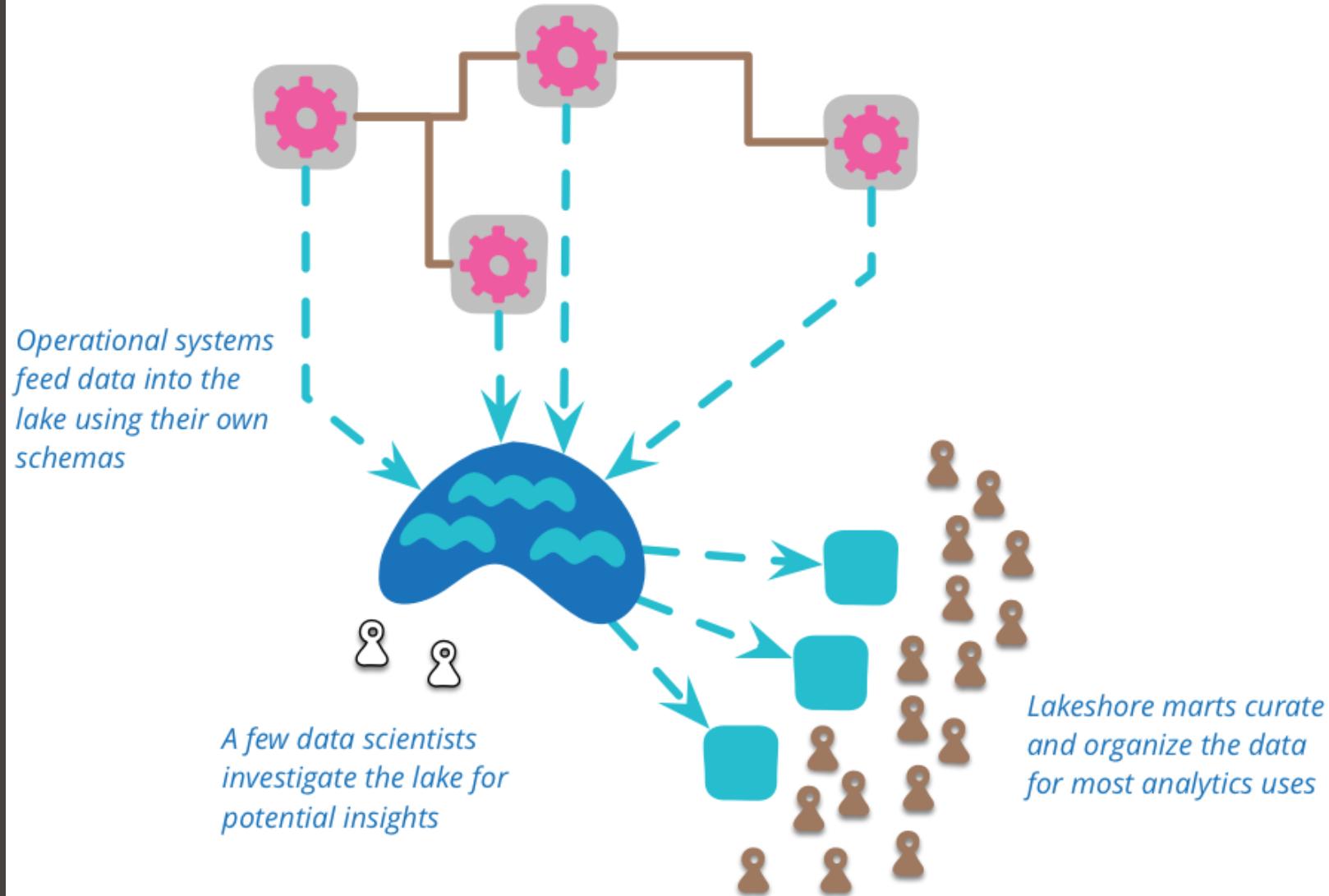
Encontrar a informação
necessária no lake pode ser
difícil?



Conceitualmente



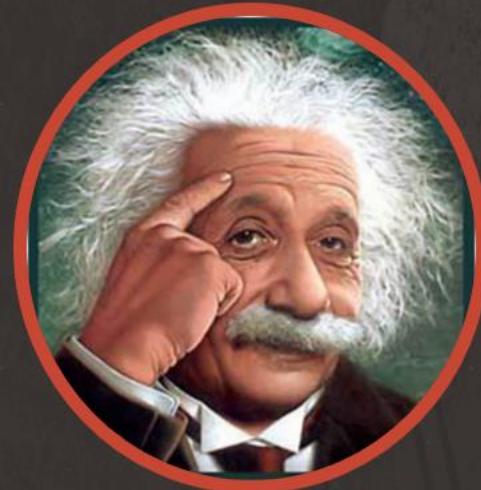
Operational systems communicate with each other via service connections, ignoring the data lake



Mas como resolver desafios na cloud?



Nós não podemos resolver **problemas** com o **mesmo pensamento** que utilizamos para **criá-los**



Albert Einstein

“
Se eu tiver **uma hora** para resolver um problema, eu gastaria **55 minutos** pensando no problema e **5 minutos** na solução

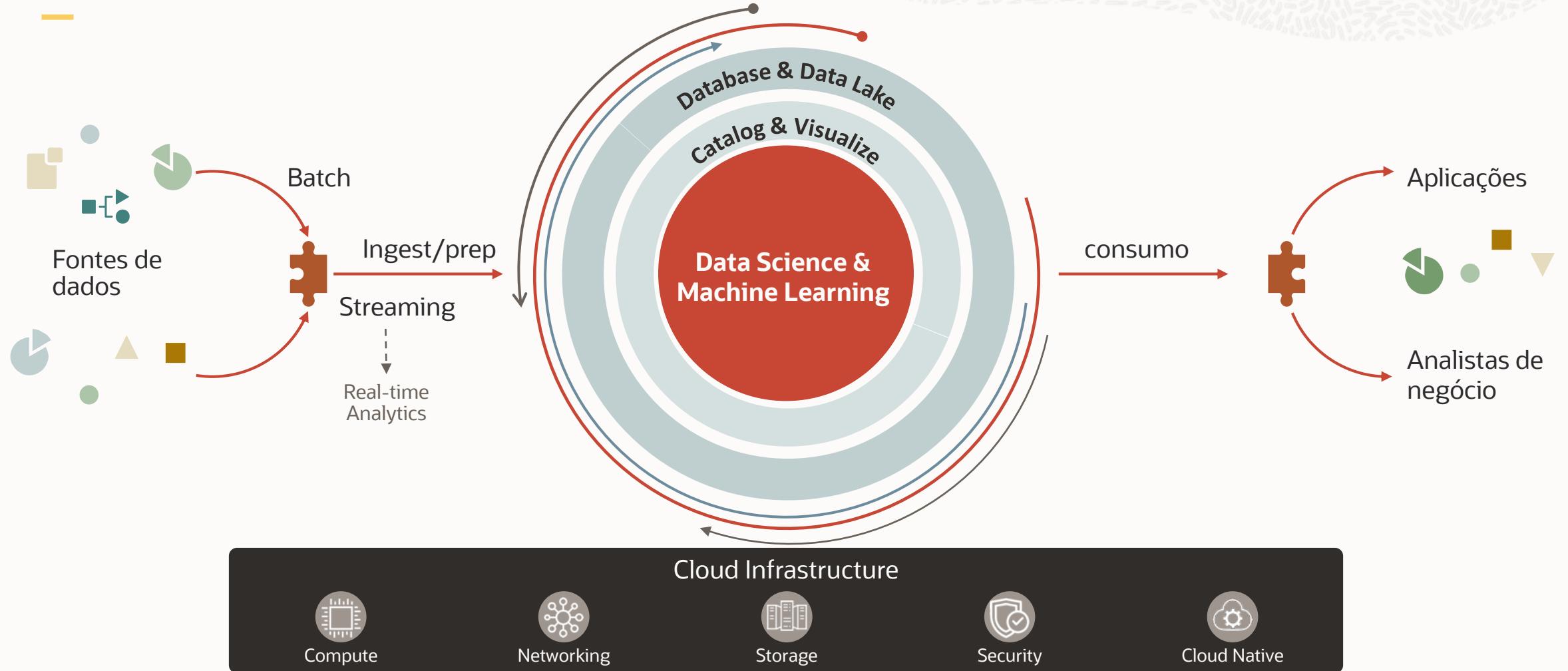
Do-it-yourself cloud

Construa o que você quer,
mas faça você mesmo



Oracle Data Platform

Machine learning suportado por dados, integração, gerenciamento e analytics



Oracle Data Platform

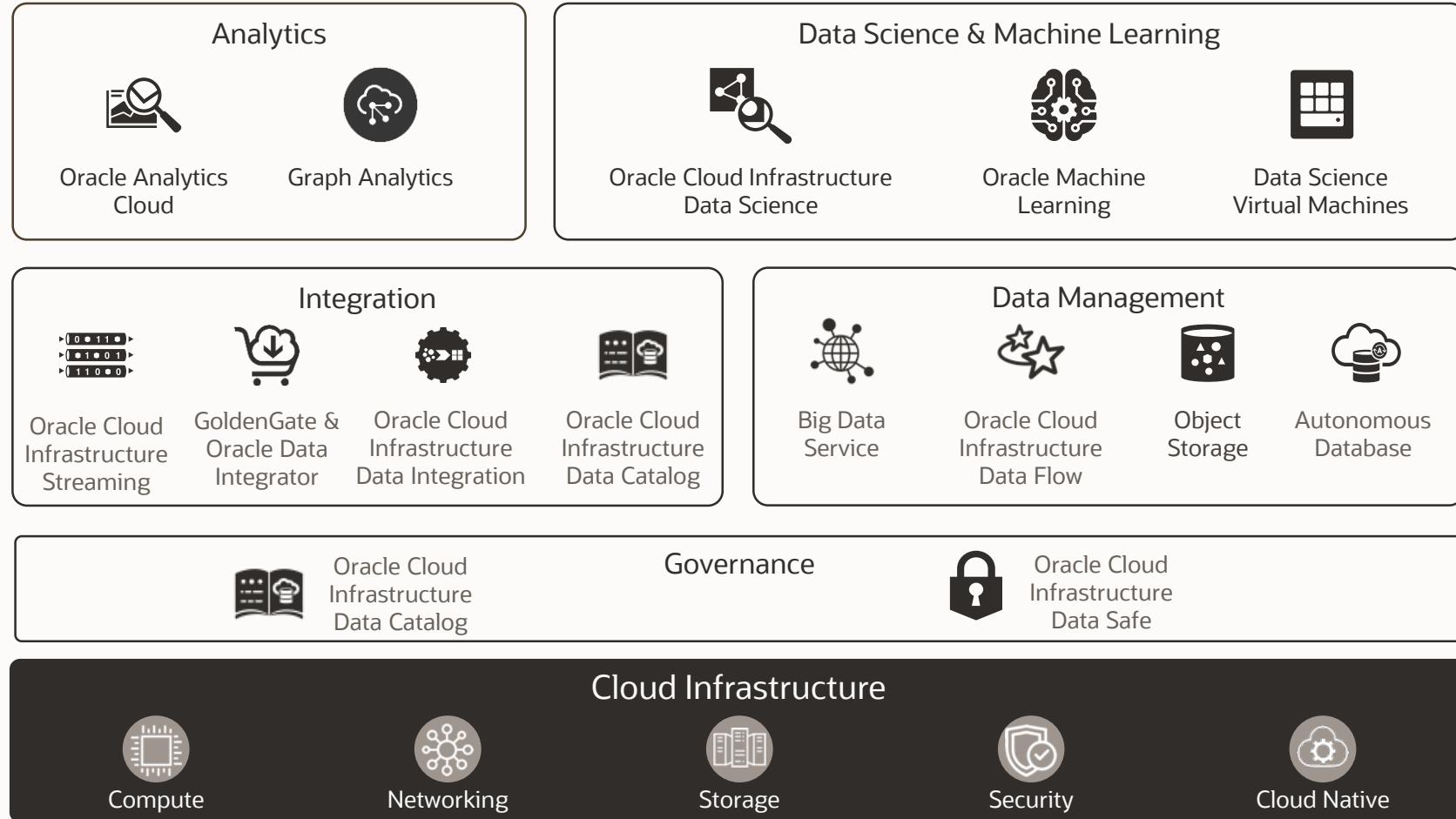
Machine learning suportado por dados, ingestão, gerenciamento e analytics



Fontes de
dados

Aplicações

Analistas de
negócio





**Mas e aí por onde começo meu
pipeline de dados?**

Jornada do Pipeline de dados

Dados

Faça um estudo de que dados você pretende armazenar, utilizar e disponibilizar em seu lake / Data Warehouse

Escolha do recurso

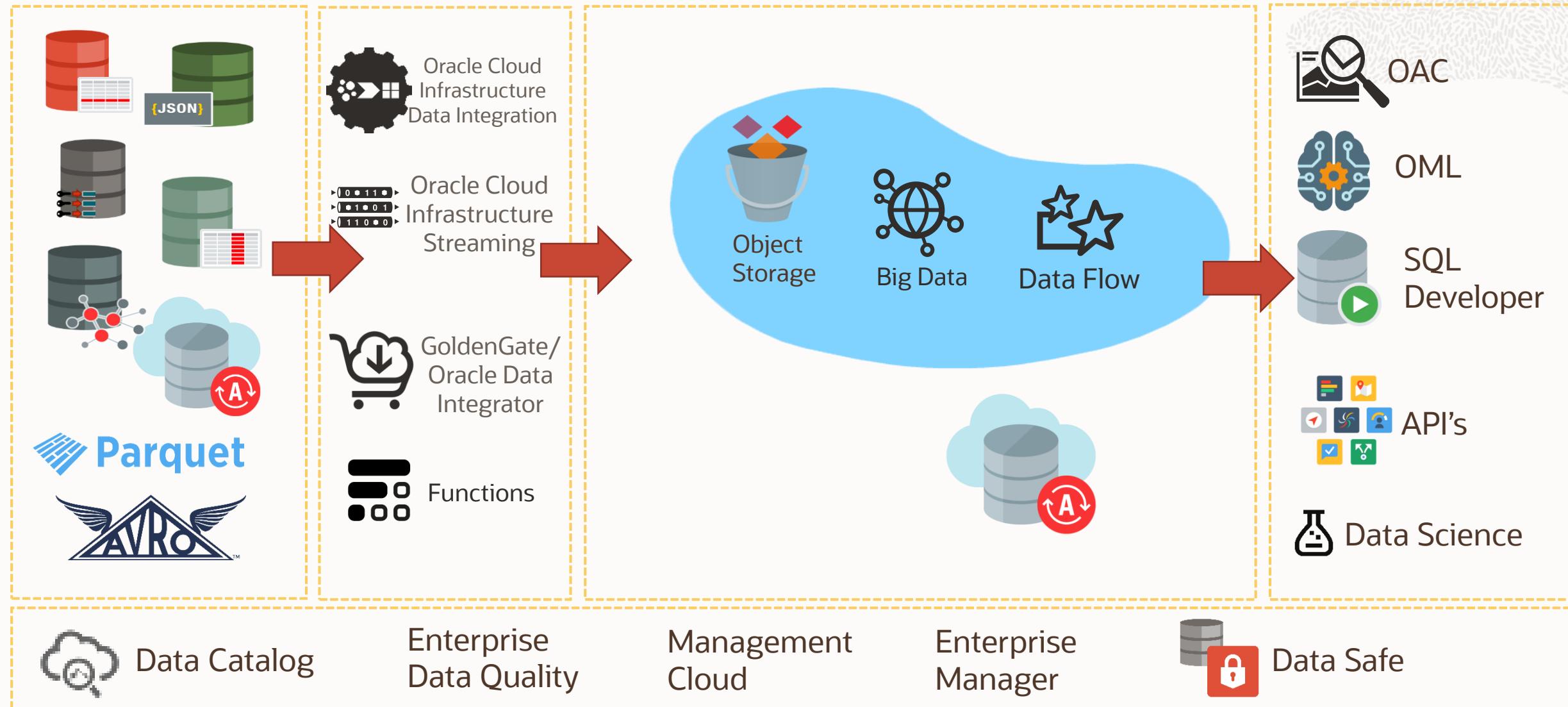
Estabelecer o melhor recurso que te atenda pode ser uma tarefa ardua, mas hoje todo provedor de cloud pode fornecer maneiras de você testar

Se questione

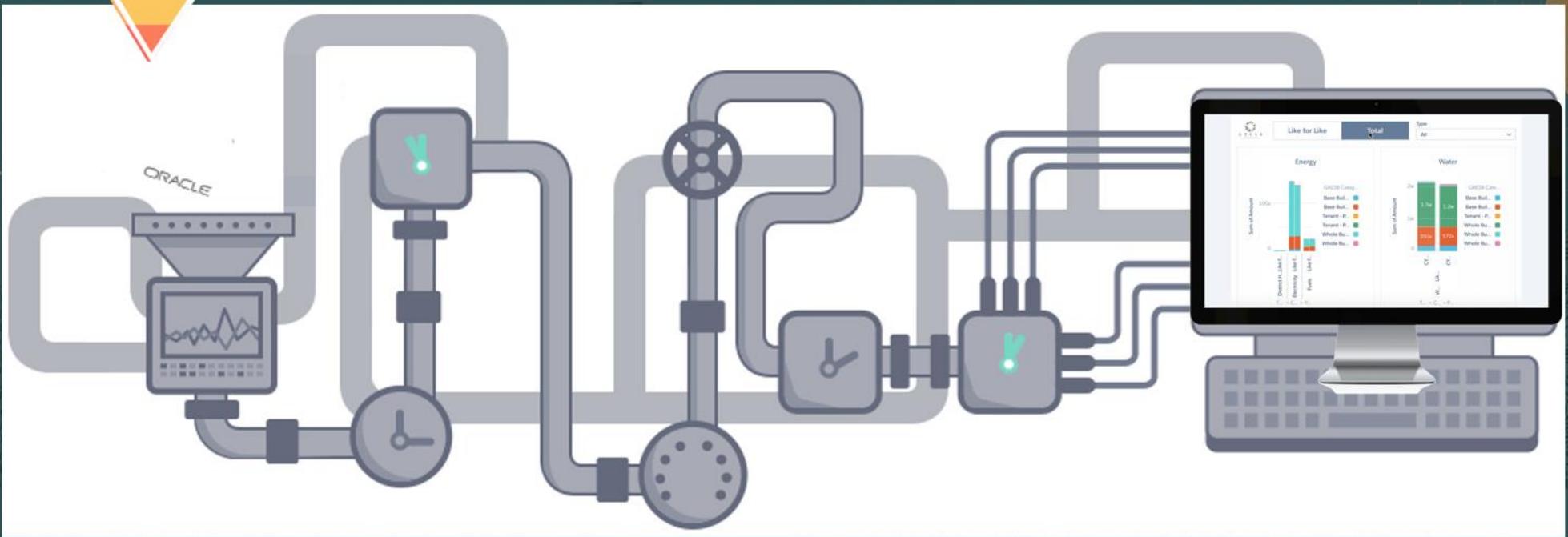
Tempo
Acesso
Ciclo de vida
Registro
Qual é o resultado final a ser esperado



Data Architecture - Reference



Show time! 😊



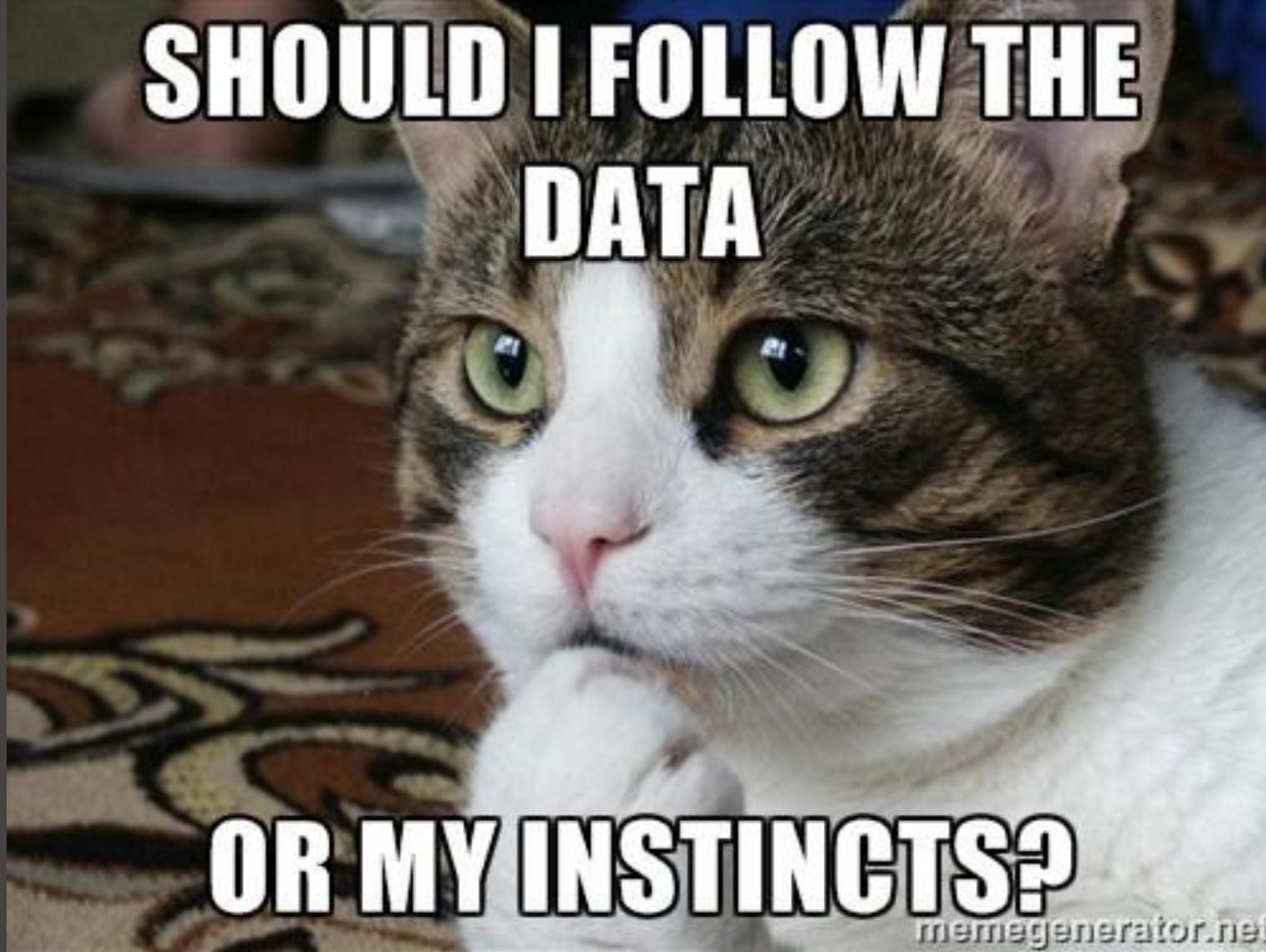


Events



Functions





**SHOULD I FOLLOW THE
DATA**

OR MY INSTINCTS?

memegenerator.net



Obrigada

Erika Nagamine

Principal Enterprise Cloud Solutions Architect

Vamos juntos nesta trilha!



linkedin.com/groups/8984009



youtube.com/c/Inovaçãocomdadosemnuvem



anchor.fm/inova-dados-nuvem



github.com/taborda-cbip/inovacao-com-dados-em-nuvem



Inovação com dados em nuvem

**CRIE SEU PIPELINE DE DADOS
NA CLOUD SEM PAGAR
NADA POR ISSO!**

