

Cloud + Dados

Será uma **combição** possível?



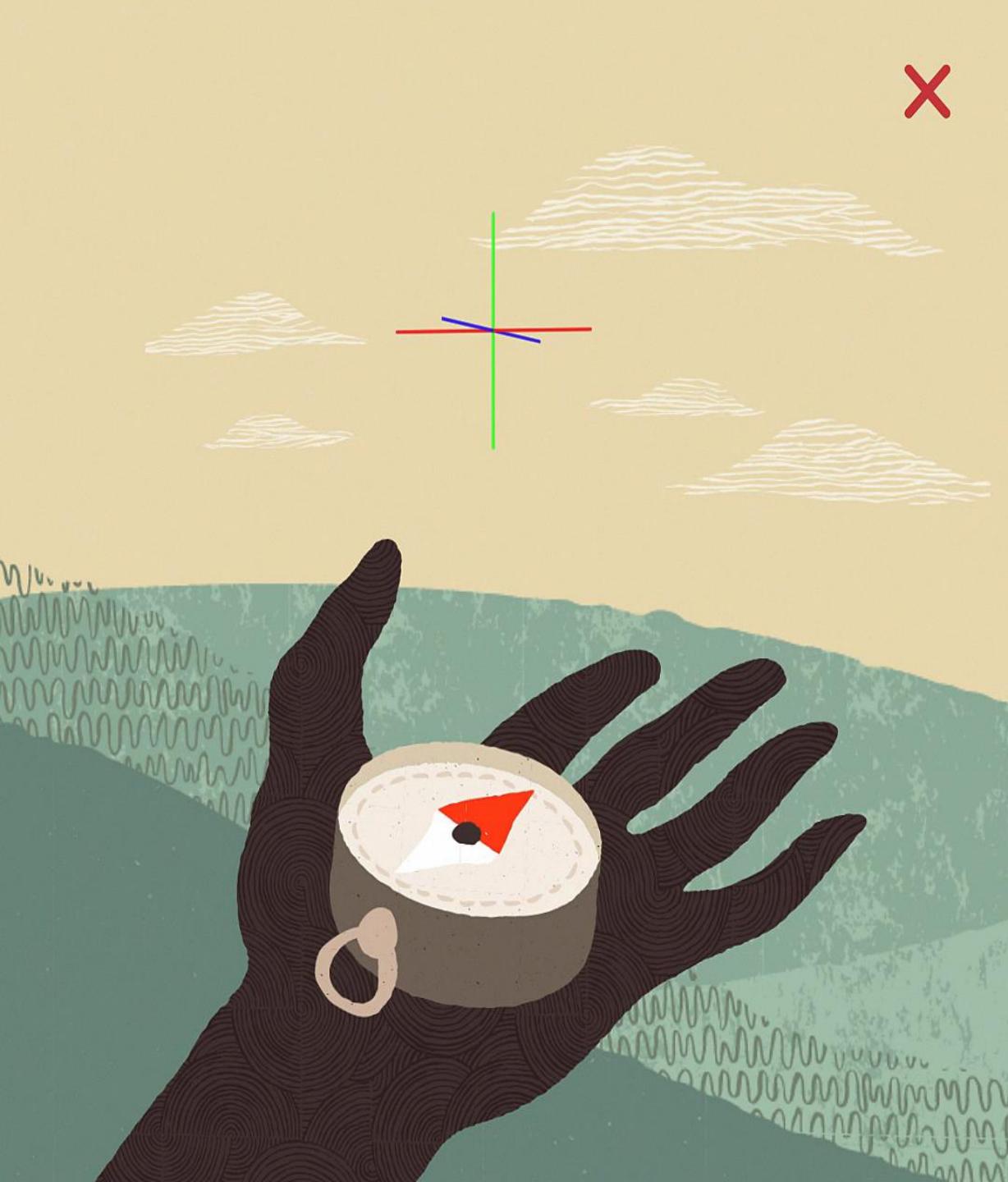
**Ciência de Dados,
Aprendizado de Máquina
e Otimização**

Erika Nagamine
28/04/2021 - 20h30
2ª CiDWeek

Safe harbor statement

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions.

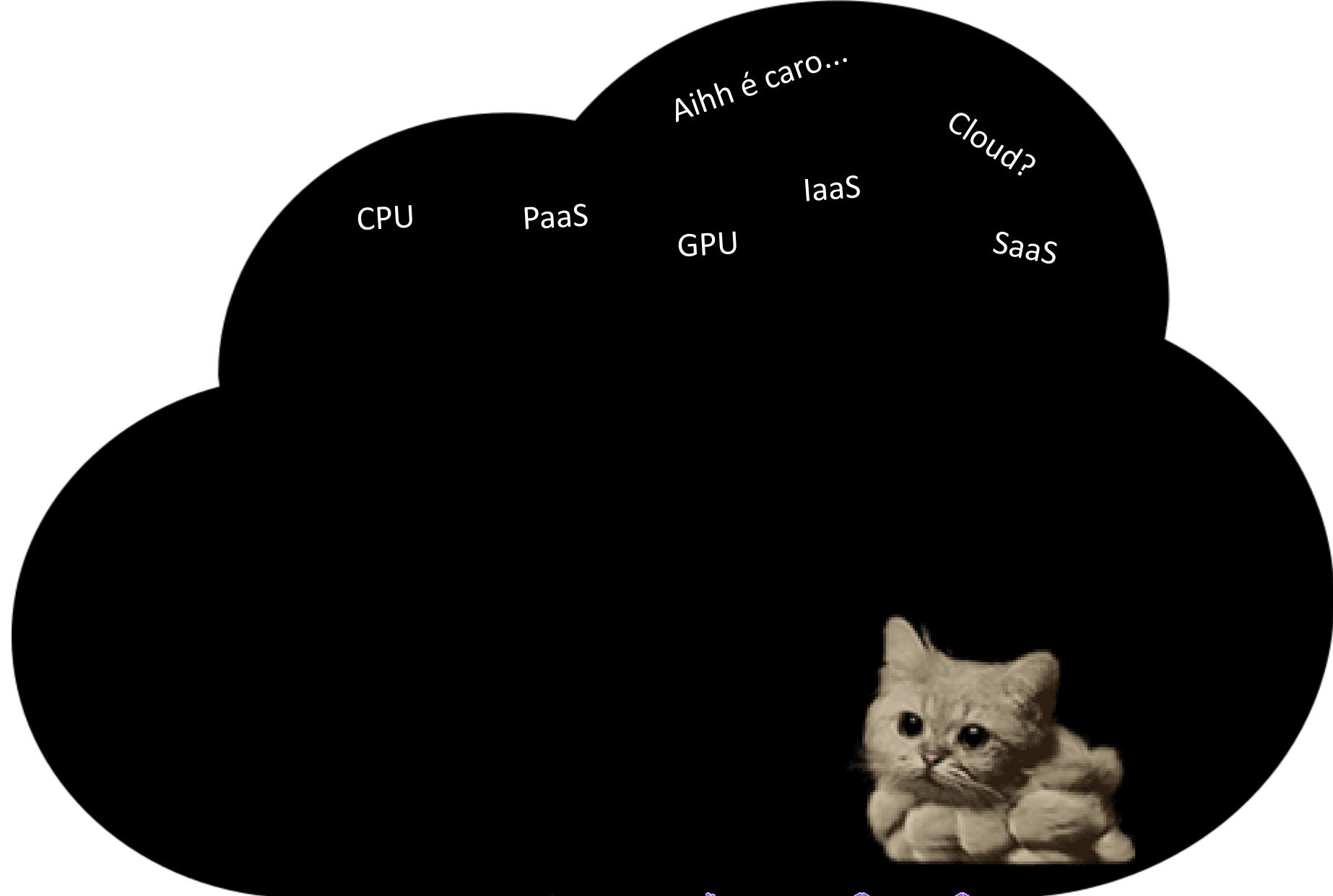
The development, release, timing, and pricing of any features or functionality described for Oracle's products may change and remains at the sole discretion of Oracle Corporation.



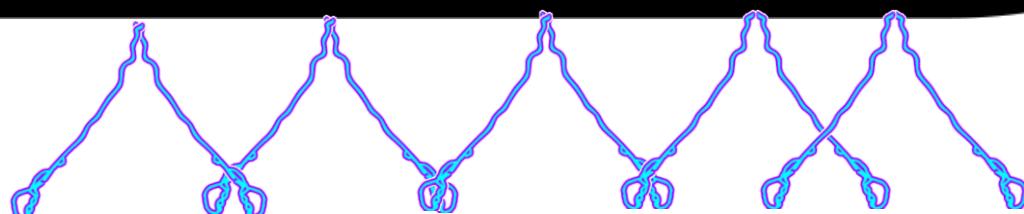
Agenda

- Contextualização
 - Dados
 - Marcos da computação
- Big Data
- Data Lake
- Projetos de Ciência de dados
- Cloud
 - Por que?
 - O que é?
 - Benefícios
 - Data Platform
 - Tipos de serviço
- Aplicabilidade de modelos de ML
 - Exemplos de modelos construídos
 - Aplicabilidade na saude
 - Detecção de distração de motorista
- Carreira
- Trilhas conhecimento



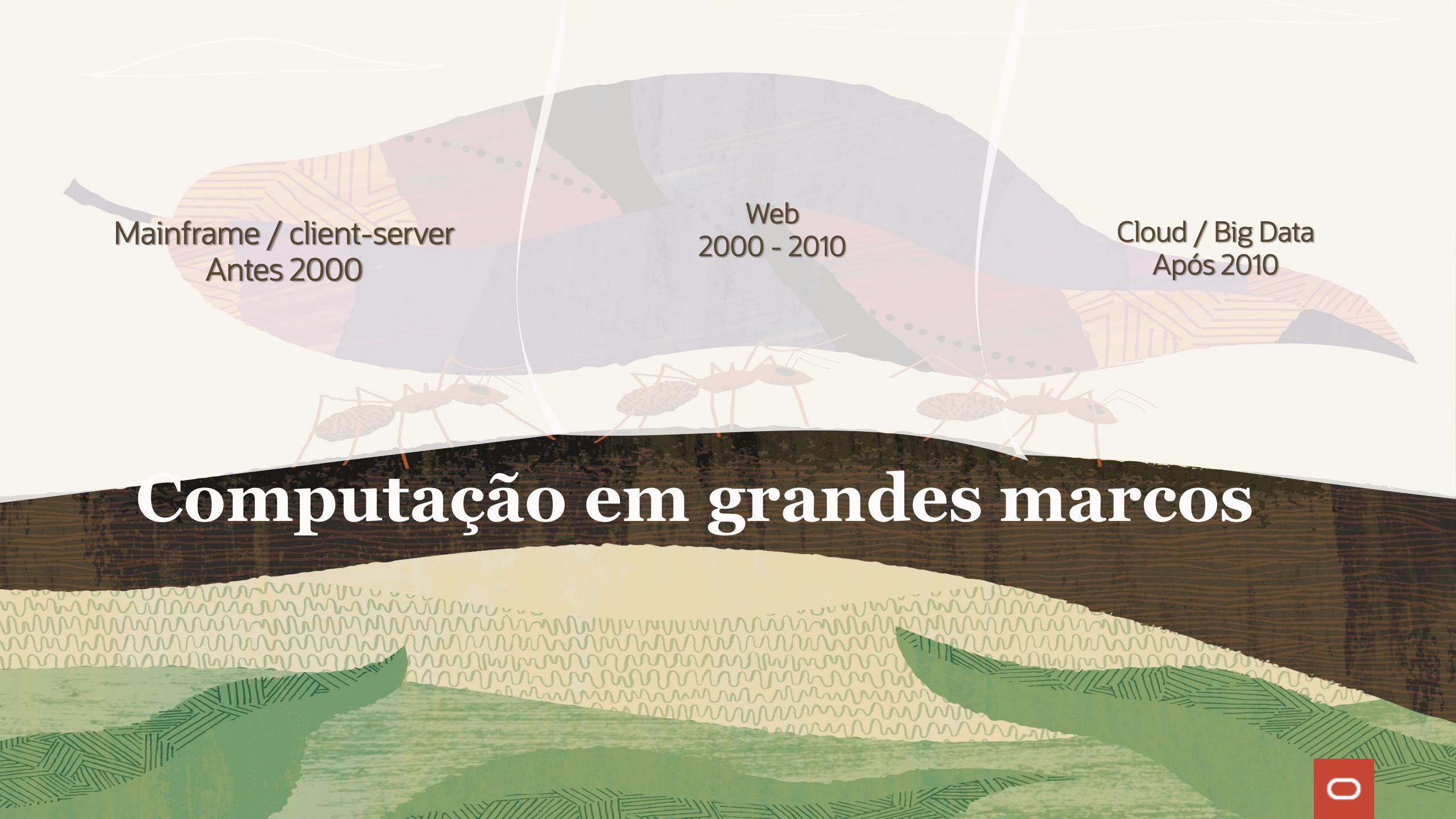


A sua solução
não pode ser
um
problema...



Dados estão mudando o mundo





Mainframe / client-server
Antes 2000

Web
2000 - 2010

Cloud / Big Data
Após 2010

Computação em grandes marcos

Big Data

Por que devemos pensar quando estamos falando de cloud?

Big Data

Value

Visualization

Variety

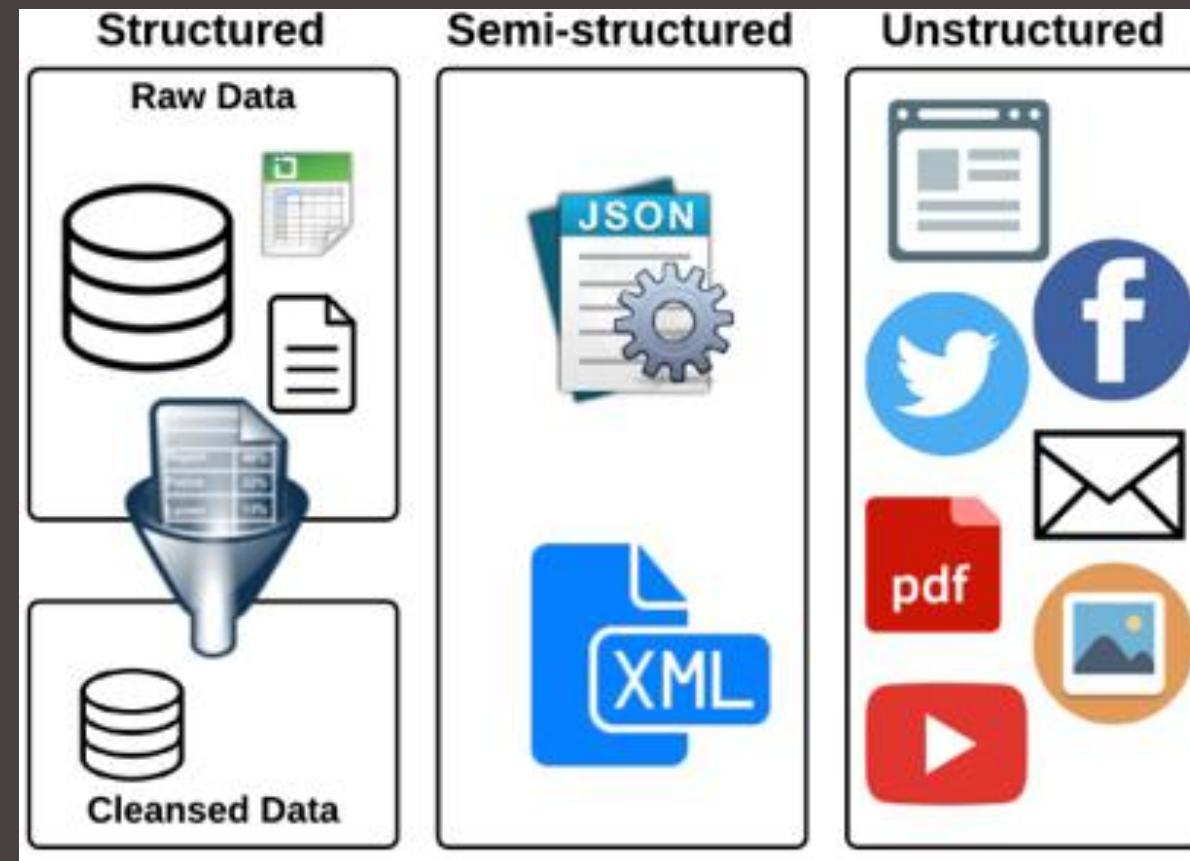
Velocity

Volume

Veracity

Vision

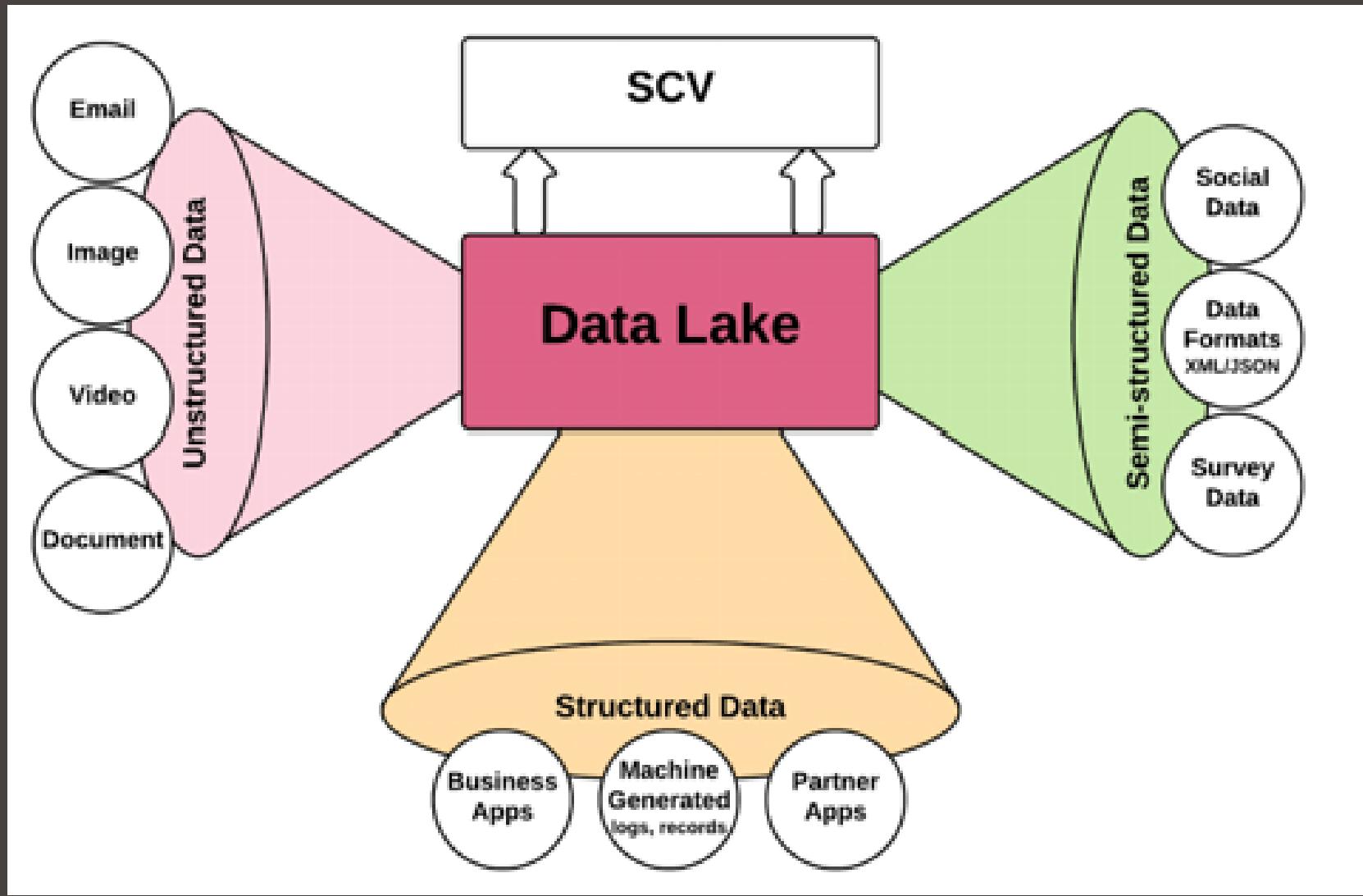
Tipos de dados



Data Lake



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



Onde gastamos mais tempo?



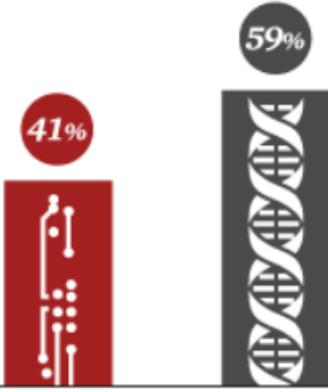
Projeto de ciência de dados



Algoritmos são melhores quando usam dados

O que não elimina o julgamento humano

Machine algorithms  Human judgment



Source: PwC's Global Data and Analytics Survey, July 2016.
Q: What will the analysis informing your next strategic decision require?
Global base: 2,106 senior executives.

Dado é ativo

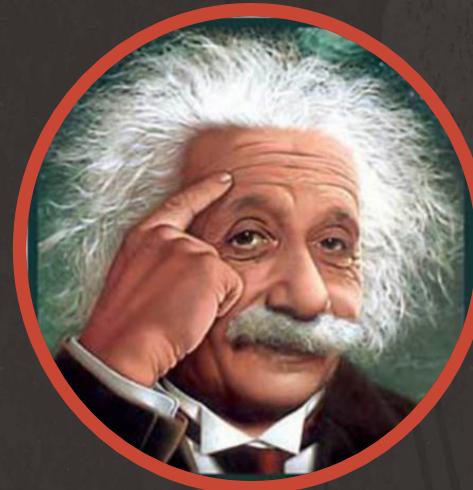


Mas como resolver desafios na cloud?



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

“



Albert Einstein

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

Do-it-yourself cloud

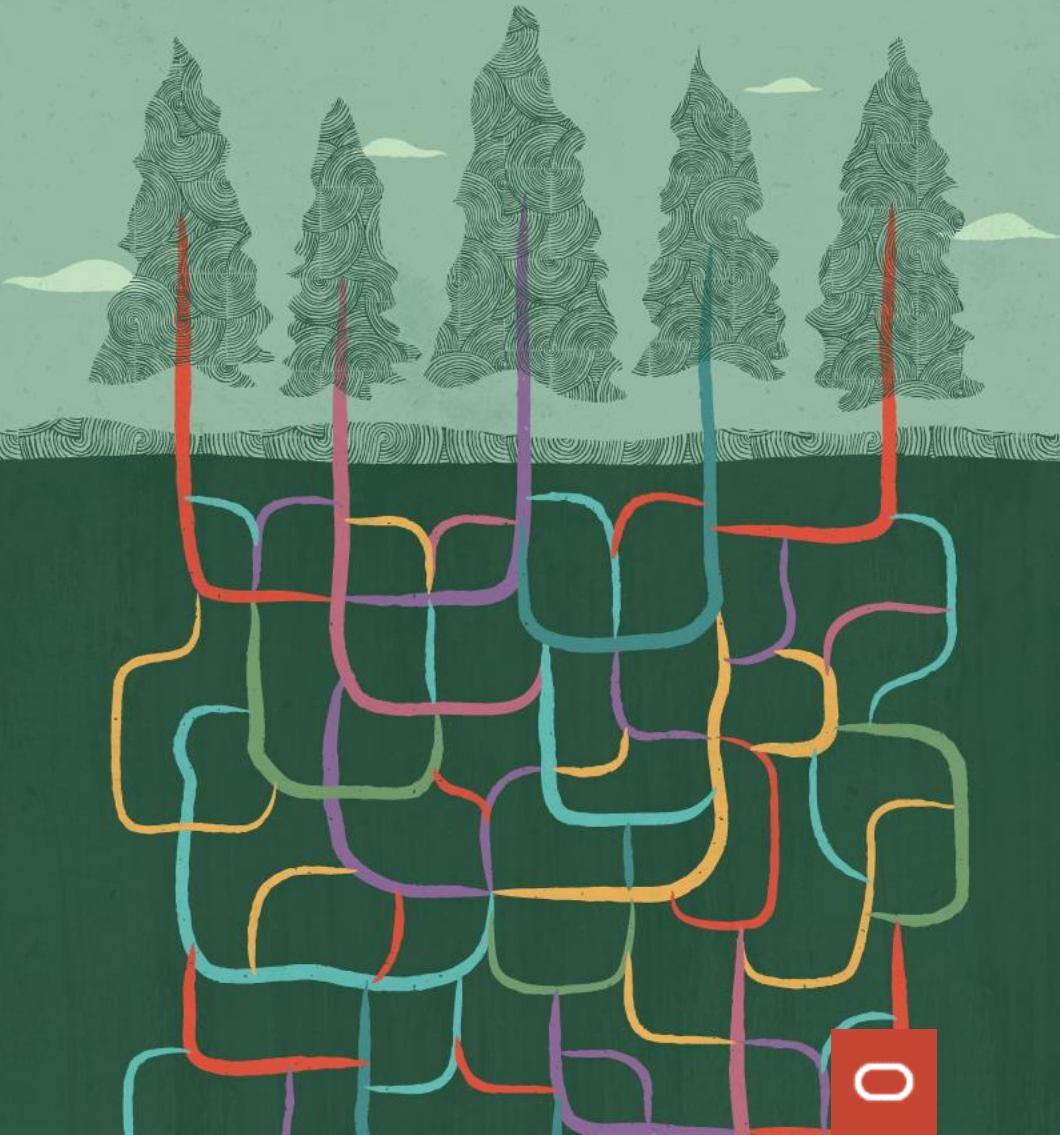
Construa o que você quer,
mas faça você mesmo



Computação em nuvem

(Cloud computing)

Entrega de recursos
computacionais sob demanda
via internet



Benefícios da computação em nuvem



Trocar custos de investimento inicial por consumo



Provisionar o que for utilizar



Presente mundialmente



Agilidade e inovação



Escalabilidade



Reducir custos de manutenção / operação de data centers



HPC, CPU e GPU

Em cloud, podemos utilizar o poder computacional para realizar tarefas incríveis



O que é High Performance Computing (HPC)

High Performance Computing (ou computação de alto desempenho) se refere a prática de **agregar poder computacional** de forma a oferecer um desempenho mais alto que um computador tradicional (desktop ou estação de trabalho comum) para resolver grandes problemas na ciencia, engenharia e negocios.

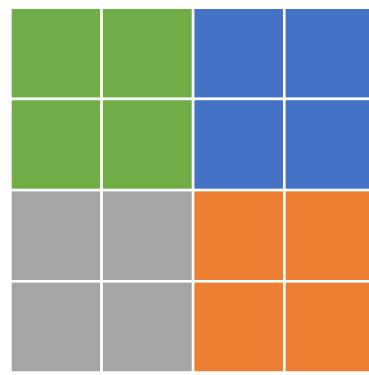
HPC na Oracle cloud pode ser oferecida como serviço:

- Bare metal CPU & GPU instances
- Fast cluster networking
- Ultralow-latency NVMe flash storage
- Block volumes & parallel file systems



Computação em paralelo

Parallel Computing – dividindo um problema para resolve-lo



- Tarefa 1
- Tarefa 2
- Tarefa 3
- Tarefa 4

- Computação em paralelo pode ser descrita como aplicar multiplos processadores para resolver um unico problema.
- Cada processador trabalha numa parte do problema
- Processadores podem trocar informações entre si.



Ingredientes

+

Preparação

+

Servir o bolo

=

Tempo
aproximado:
3 Horas



Ingredientes

Supermercado:
4 pessoas



Preparação

bolo: 1 pessoa
Recheio: 1 pessoa
Composição do bolo completo: não paralelizada



Servir o bolo

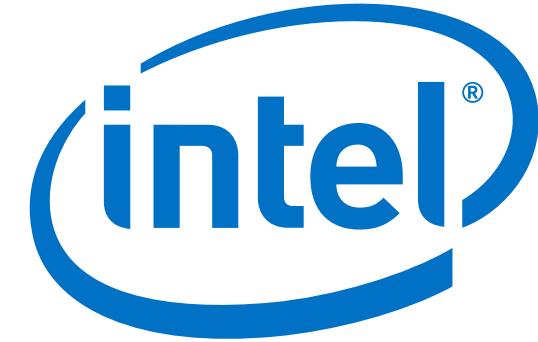


Tempo
aproximado:
2 Horas



CPU x GPU

- **CPU:** para cargas de trabalho pesadas, projetada para resolução de lógicas com controle complexos e é otimizado para programas sequenciais.
- **GPU:** para cargas de trabalho leves, desenhadas para resolução de problemas com lógicas com controle simples, otimizadas para tarefas com paralelismo e grande foco em alto desempenho de programas em paralelo,
- **CPU+GPU:** Arquitetura heterogênea



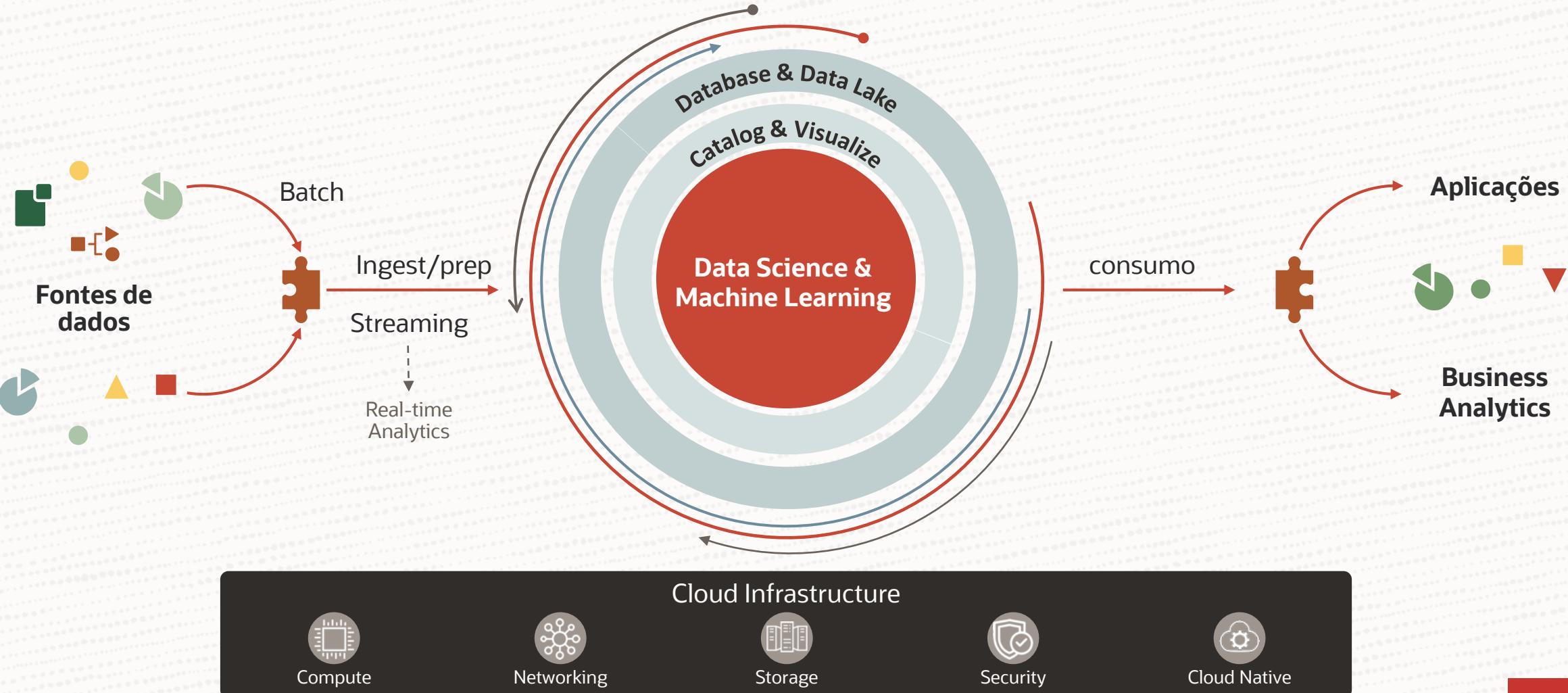
Cloud + Dados



Existem diversas formas de resolver o mesmo problema, das mais simples às mais complexas

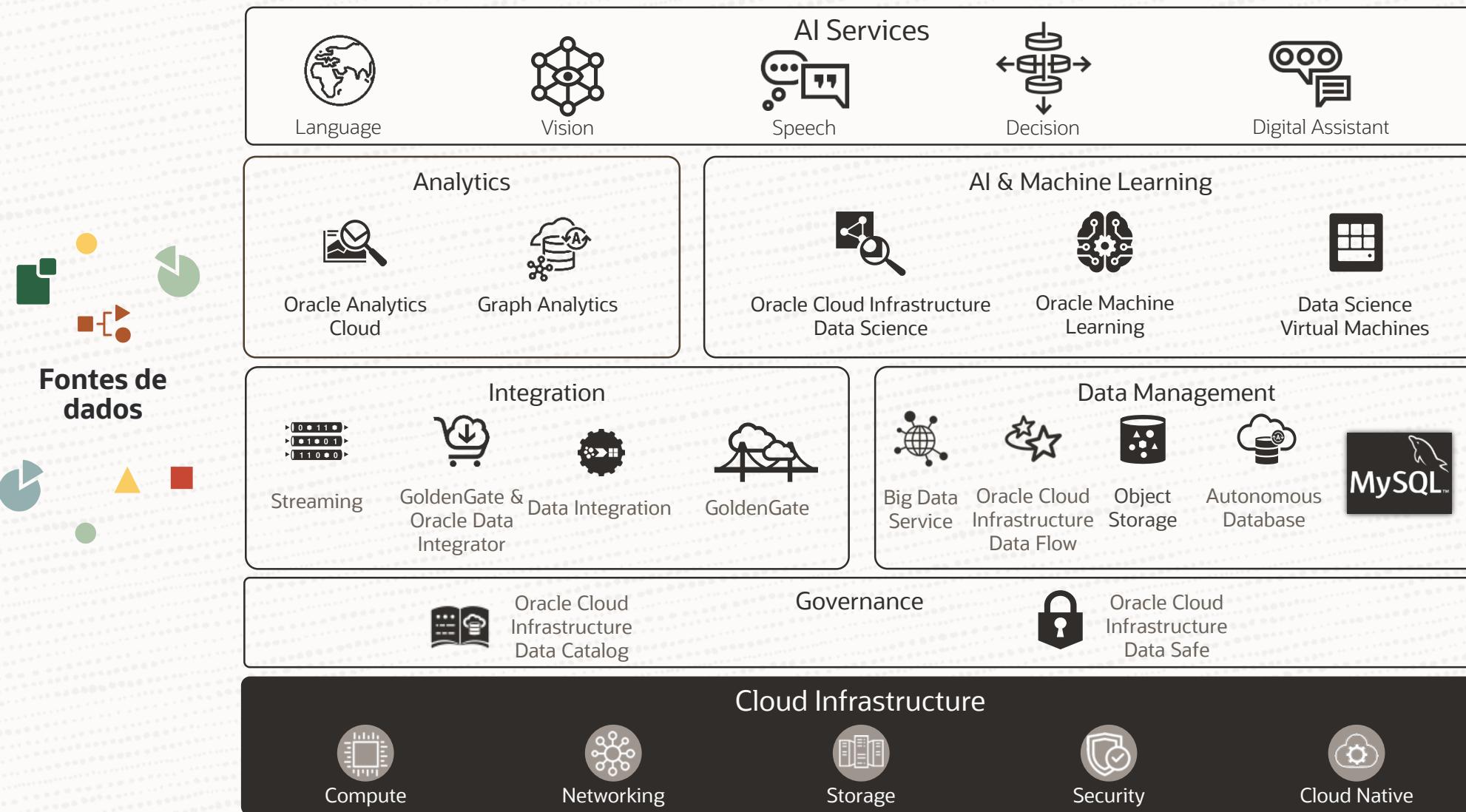
Oracle Cloud Data Platform

Machine learning suportado por dados, ingestão, integração, gerenciamento e analytics



Oracle Cloud Data Platform

Machine learning suportado por dados, ingestão, integração, gerenciamento e analytics



Que serviço escolho para iniciar meu mundo em cloud



Tempo



PaaS
SaaS

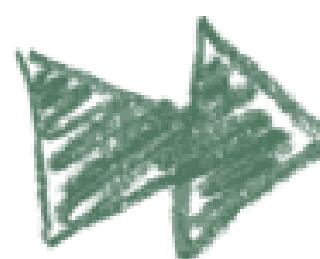


IaaS

Maior Agilidade,
foco no produto
não tecnologia

Habilidades
específicas são
necessárias

Esforço



Necessidade
conhecimento

Solução



Mas e aí por onde começo minha
Jornada Cloud com Dados?

Jornada Cloud + Dados

Dados

Faça um estudo de que dados você pretende armazenar, utilizar e disponibilizar

Escolha do recurso

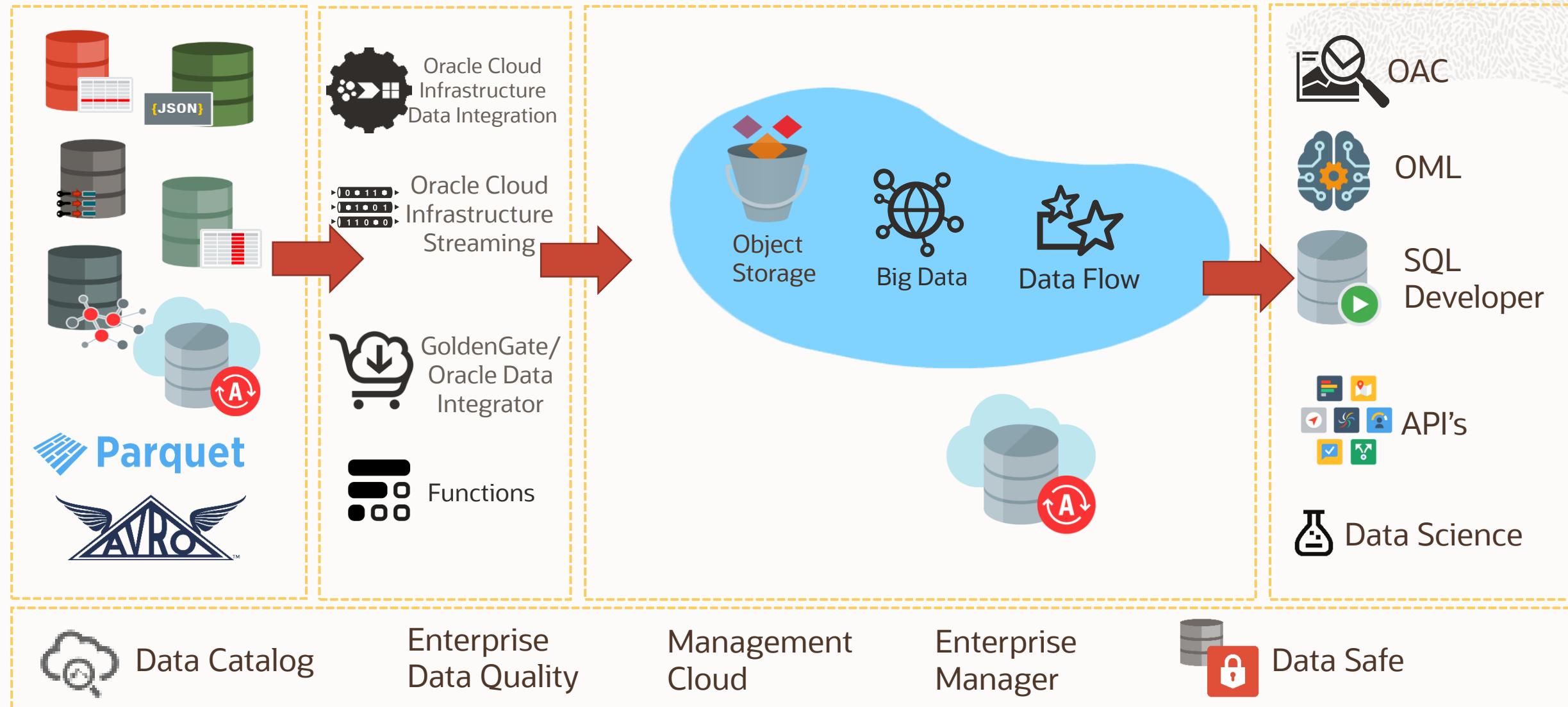
Estabelecer o melhor recurso que te atenda pode ser uma tarefa árdua, mas hoje todo provedor de cloud pode fornecer maneiras de você testar

Se questione

Tempo
Acesso
Ciclo de vida
Registro

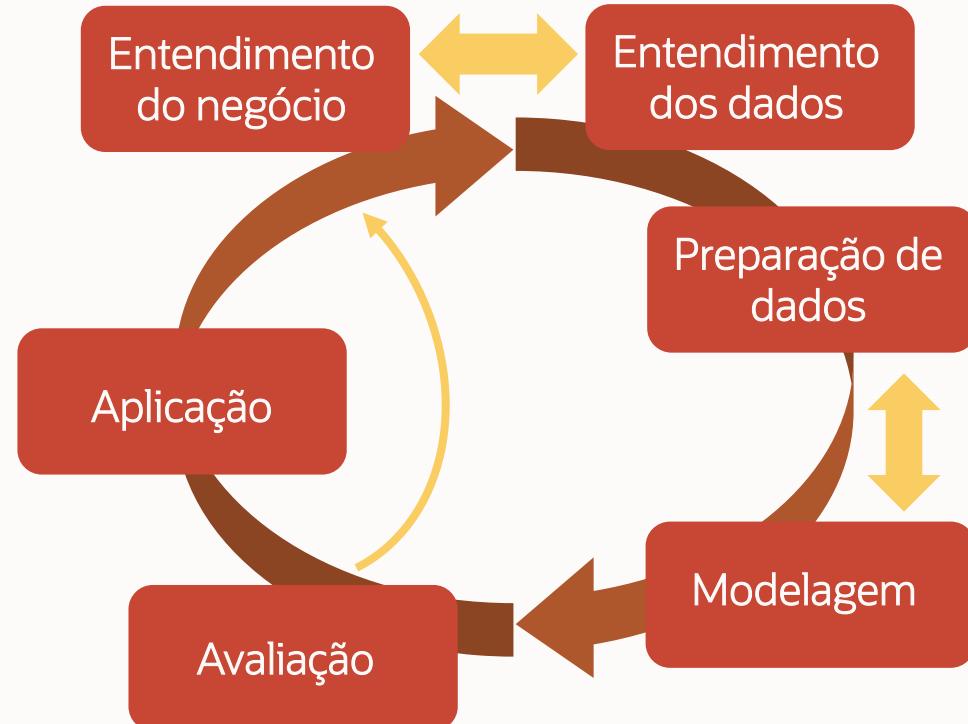


Data Architecture - Reference



Construção de algoritmos de ML

Metodologia CRISP-DM



Cross-industry standard process for data mining

Oportunidades de automação

Preparação de dados*
Modelagem
Avaliação
Aplicação

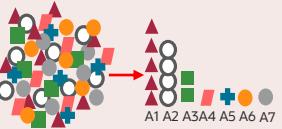
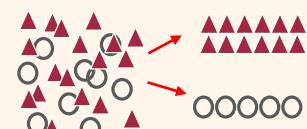
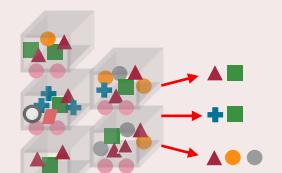
Desafios de automação

Preparação de dados*
Entendimento do negócio
Entendimento dos dados

*Depende do tipo de dados. Tendências baseadas no tempo são um desafio, a criação de característica é difícil sem interpretação

Aplicabilidades de Machine Learning

Algoritmos *automaticamente* filtram grandes quantidades de dados para descobrir padrões, extrair novas idéias e fazer previsões

| Quais são atributos que explicam o padrão de compra dos clientes? | Identificação / predição dos melhores clientes | Previsão de vendas / pedidos nos meses futuros | Identifica o melhor cliente usando modelo RFM (Recency, Frequency, Monetary) | Identifica possíveis atividades fraudulentas | Sugere itens adicionais ao cliente baseado no consumo |
|---|--|---|--|--|--|
| Attribute Importance | Classification | Regression | Clustering | Anomaly Detection | Associations |
|  |  |  |  |  |  |
| Identifica o fator mais importante que explica o evento | Prediz o sentimento do cliente e encontra padrões de consumo | Prediz ou estima um valor | Segmenta uma população em silos | Encontra outlier or “eventos raros” | Determina a ocorrência similar de consumo |

Outro exemplo prático de como ML podem ajudar no dia a dia



A triagem automática do melanoma é uma tarefa desafiadora pois a diferença entre a lesão cancerosa (linha superior) e a não cancerosa (linha inferior) muitas vezes é sutil

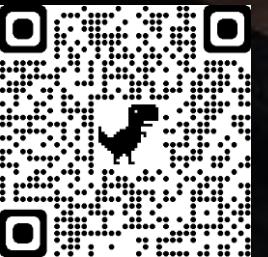
Distracted Driver Detection

Entrega final

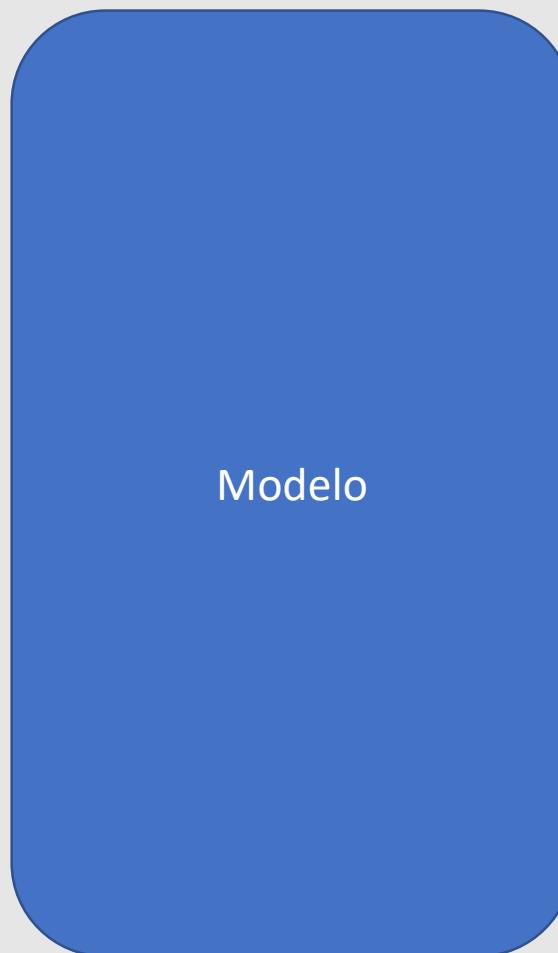
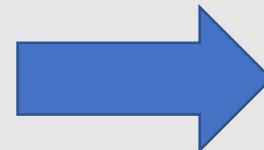
Alexandre Gomes
Cristofer Girardi
Erika Nagamine
Gabriel Machado
Henrique Capatto

UNICAMP

Mineração de dados complexos



Detecção de distração do motorista



| Classes |
|---------|
| C0 |
| C1 |
| C2 |
| ... |
| ... |
| C9 |

$$\text{logloss} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{ij} \log(p_{ij}),$$

multi-class logarithmic loss

Exceções



Padrão



Classe c9: Falando com o passageiro

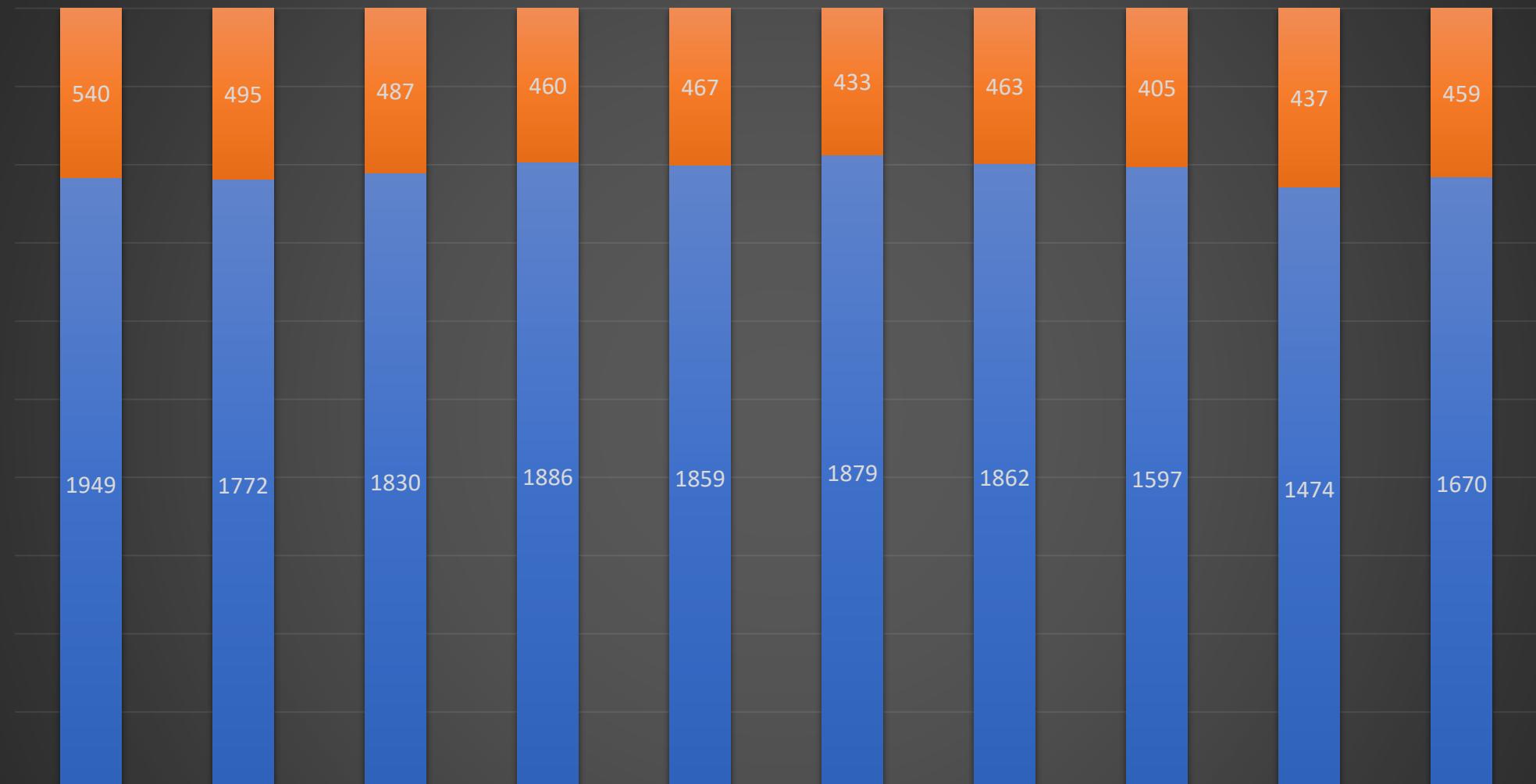
Padrão observado

- Pessoas sentadas com a cabeça virada e boca levemente aberta

Exceções

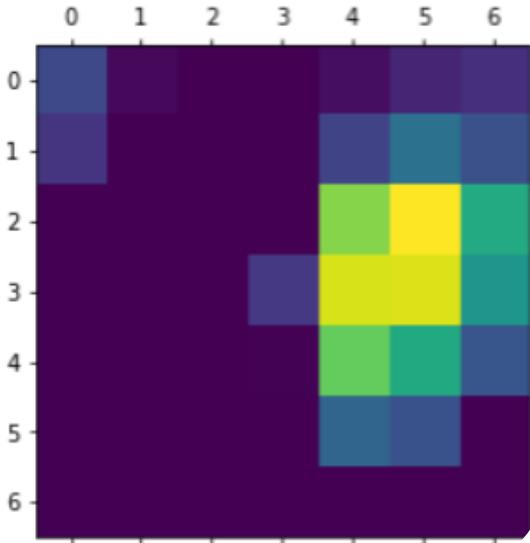
- Mão fora do volante
- Pessoas com mão no volante mas olhando para o lado

Divisão das classes - Baseline

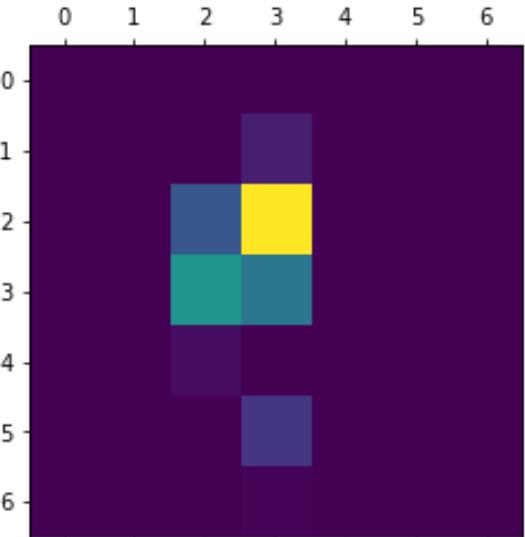


| | c0 | c1 | c2 | c3 | c4 | c5 | c6 | c7 | c8 | c9 |
|-----------|------|------|------|------|------|------|------|------|------|------|
| validação | 540 | 495 | 487 | 460 | 467 | 433 | 463 | 405 | 437 | 459 |
| treino | 1949 | 1772 | 1830 | 1886 | 1859 | 1879 | 1862 | 1597 | 1474 | 1670 |

**c9: Falando com o passageiro
(correto) x c0: direção segura
(predito)**



**c9: Falando com
passageiro (predito
corretamente)**



Resultados obtidos: VGG16
(modelo)

Entendimento dos resultados

Um pouco sobre a solução



Aplicabilidade prática



YouTube



Distracted
Driver
Detection

Grupo Outliers

Alexandre Gomes
Cristofer Girardi
Erika Nagamine
Gabriel Machado
Henrique Capatto

UNICAMP
Mineração de dados complexos
MDC-007

Fonte: <https://mitchellwhale.com/blog/distracted-drivers/>



Encontrar a carreira ideal
que desejo seguir pode ser
um passo difícil...

Dados + Cloud



Carreiras possíveis

DBA

Engenheiro
de dados

Cientista
de
dados

Engenheiro
de
Machine
Learning

Analista
de
Dados /
Analytics

Arquiteto
de dados

Engenheiro
de Cloud

Devops /
SRE

Arquiteto
Cloud

Entre outras...

Quero obter +
conhecimento



Formas de conhecimento

Site: <https://apexapps.oracle.com/pls/apex/dbpm/r/livelabs/home>

Welcome to Oracle LiveLabs!

About LiveLabs: Oracle LiveLabs gives you access to Oracle's tools and technologies to run a wide variety of labs and workshops. Experience Oracle's best technology, live!

Shortcuts: View All Workshops | Try an Always Free Workshop

Featured Workshops:

- Autonomous Database Quick Start
- Get Started with Oracle Database
- DB Security Basics
- Converged Database for Developers
- Migrate WebLogic Server to OCI with Autonomous Database

User icons: Application Developer, Low Code Developer, Data Scientist, DBA, DevOps Engineer

Site: <https://education.oracle.com/learning-explorer>

ORACLE LEARNING EXPLORER

FREE TRAINING & ACCREDITATION

Free Training Content: Get started with free Oracle Cloud training in our most popular areas. Develop in-demand skills employers need.

Start learning for free

Software as a Service (SaaS): Human Capital Management, Talent Management, Enterprise Resource Planning, Supply Chain Management, Procurement, CX Sales, CX Configure, Price, and Quote, CX Marketing, Customer Experience, Enterprise Performance Management.

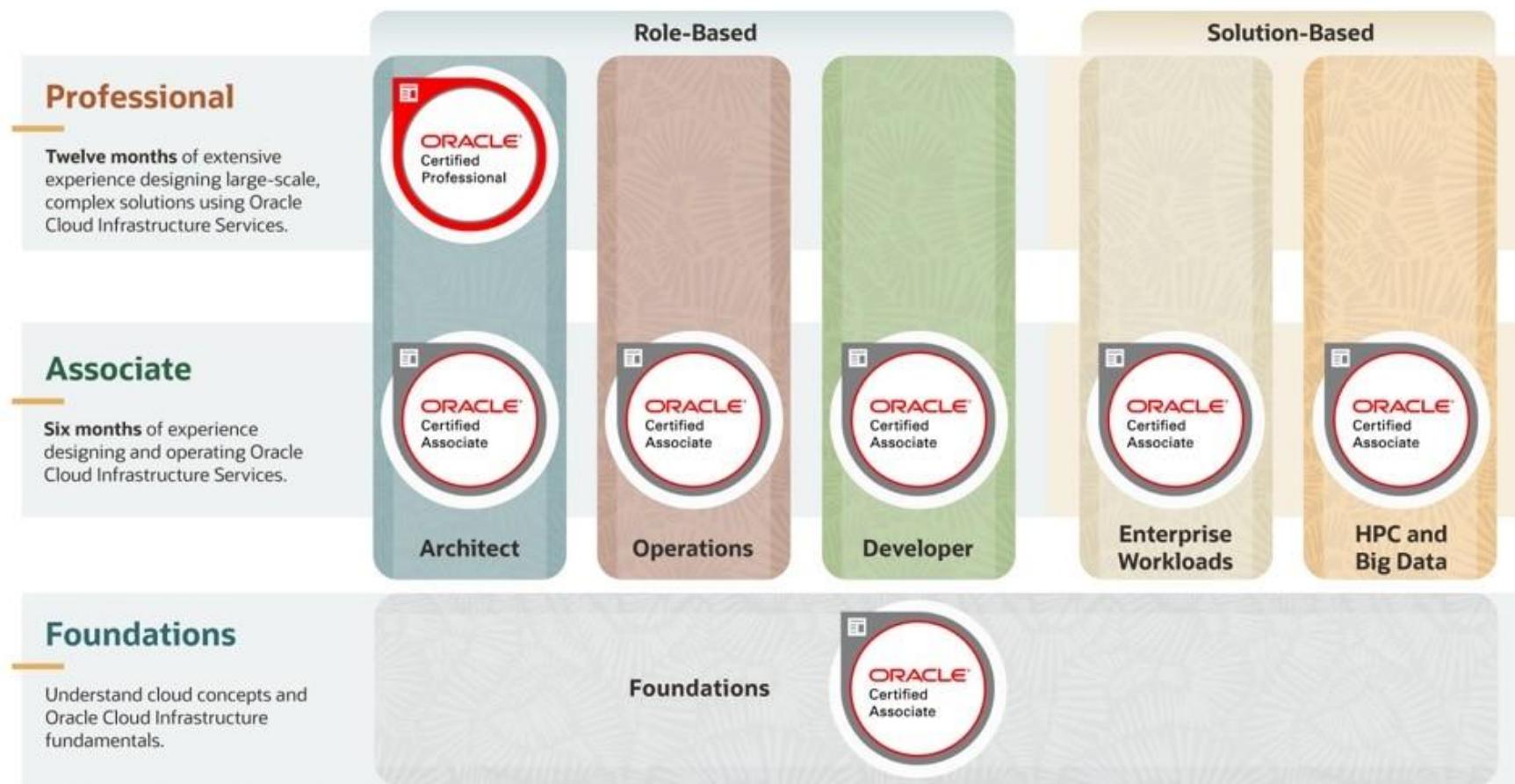
Oracle Cloud Infrastructure (OCI): OCI.

Platform as a Service (PaaS): Database Management, Application Development, Oracle Database, Java, MySQL.

ORACLE LEARNING EXPLORER

O

Certificações



Site: https://education.oracle.com/oracle-certification-path/product_683



Always Free

Serviços grátis
para sempre



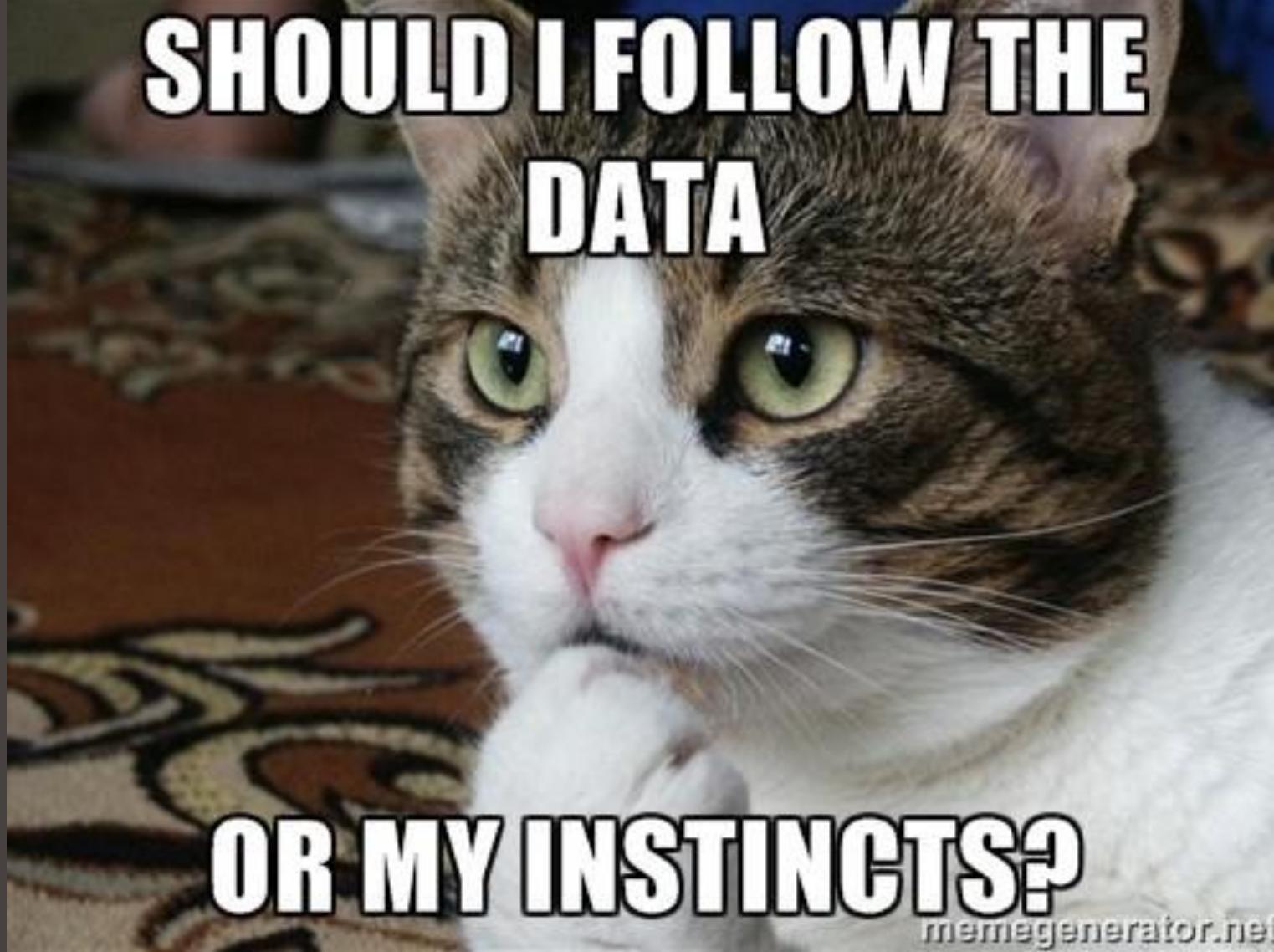
Free Trial

Créditos gratuitos



<http://oracle.com/br/cloud/free>





**SHOULD I FOLLOW THE
DATA**

OR MY INSTINCTS?

memegenerator.net



Obrigada

Erika Nagamine

Principal Enterprise
Cloud Solutions Architect

