



# Machine Learning e IA em Ambientes Distribuídos







Data Science  
Academy

Data Science Academy eng.davidborges@gmail.com 59532d8f5e4cdead748b456a

Bem vindo(a)!

# Machine Learning e IA em Ambientes Distribuídos

**Esse é o quarto curso da Formação Engenheiro de Dados**



Data Science Academy



# APRESENTAÇÃO



Introdução



Instrutores



DSA



Módulos



Metodologia

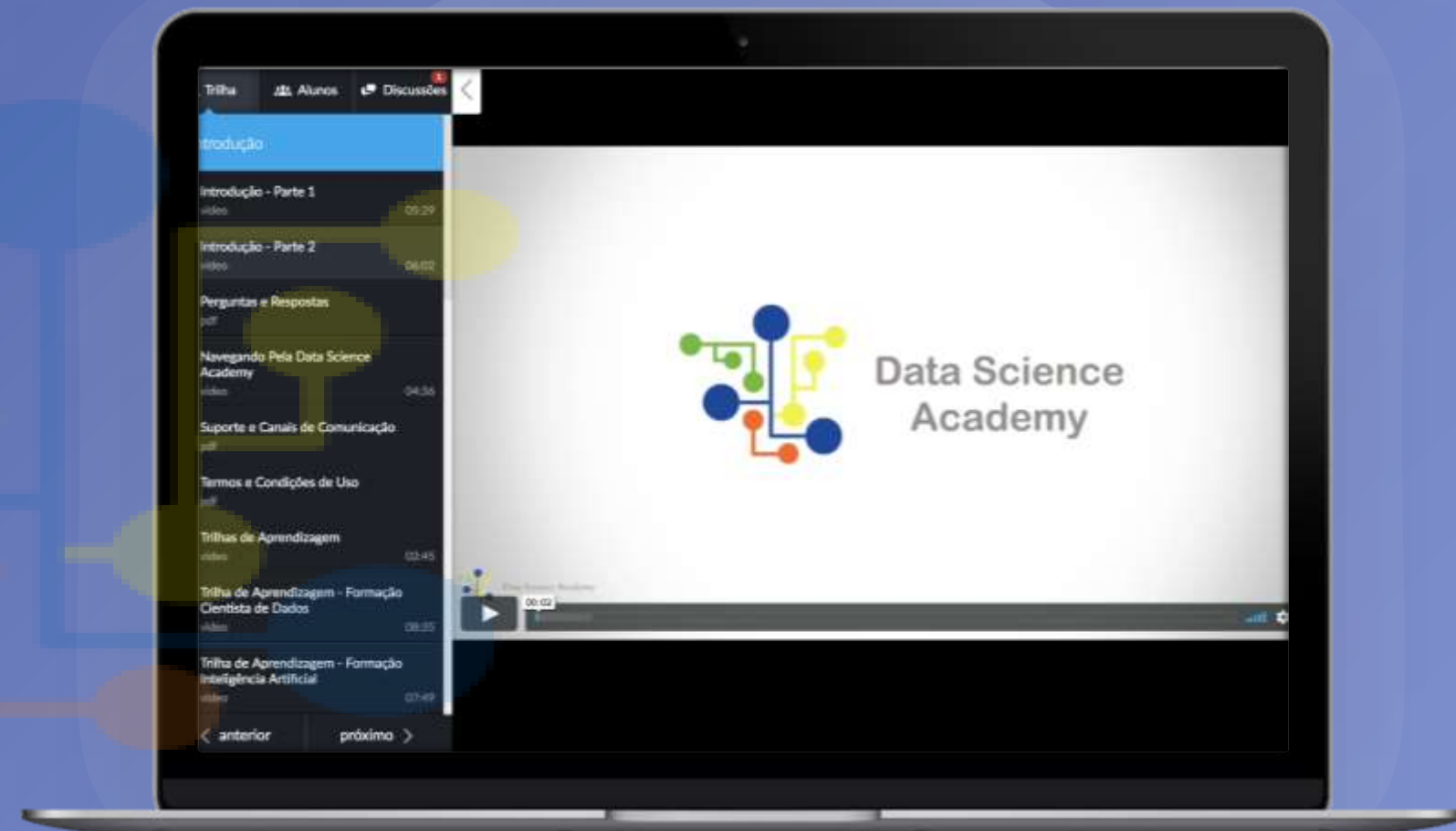


Avaliação



# Nossa Academia

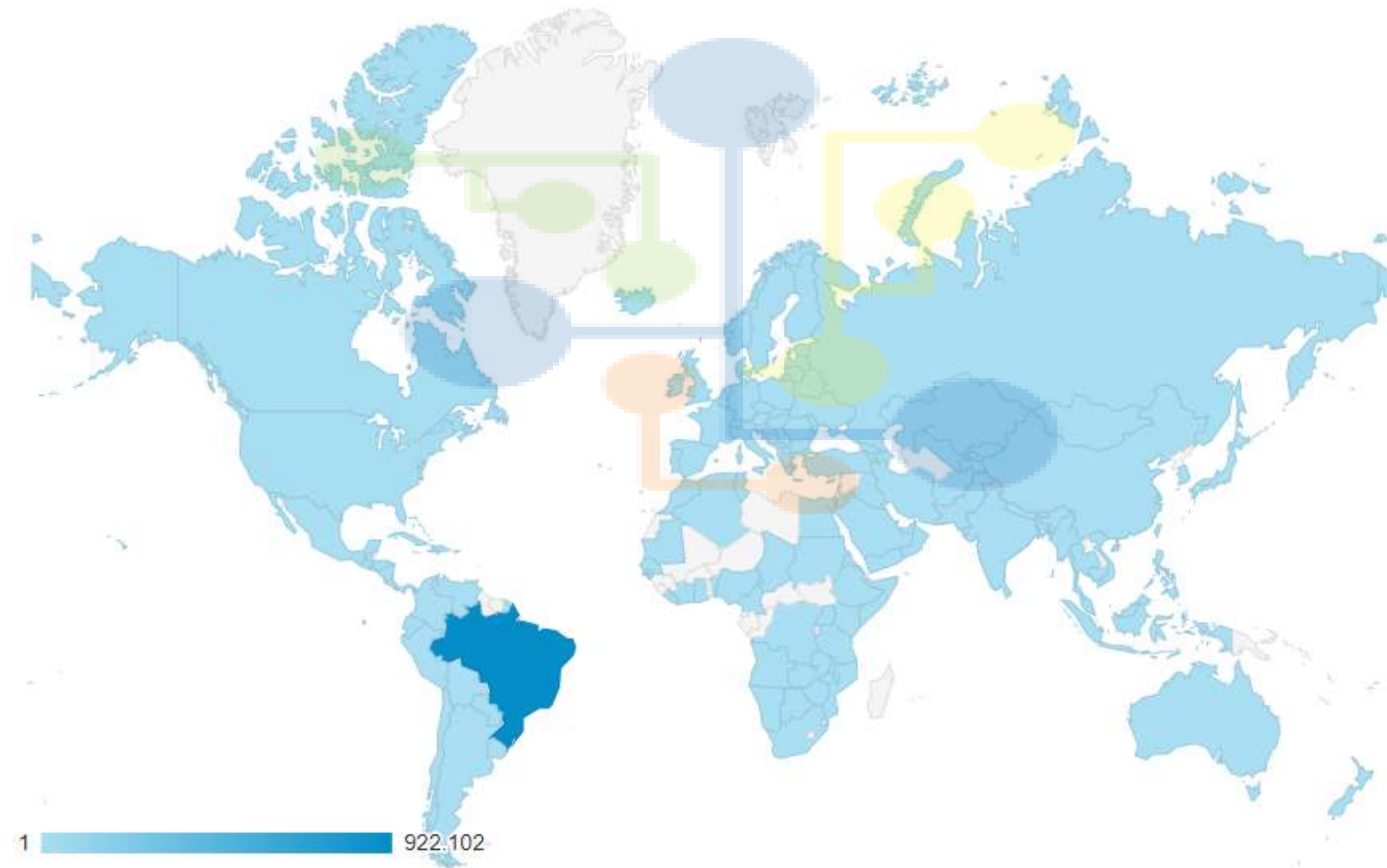
A Data Science Academy (DSA) é um portal de ensino online especializado em Big Data, Machine Learning, Inteligência Artificial, Desenvolvimento de Chatbots e tecnologias relacionadas. Nosso objetivo é fornecer aos alunos conteúdo de alto nível por meio do uso de computador, tablet ou smartphone, em qualquer lugar, a qualquer hora, 100% online e 100% em português.





# Data Science Academy - Localização

No Brasil e no Mundo





# Curso técnico, voltado para profissionais de tecnologia.

## Pré-requisitos

Esperamos que você tenha um bom conhecimento em Sistemas Operacionais (especialmente sistema Linux) e a Formação Engenheiro de Dados oferece como bônus um curso completo de Introdução ao Sistema Operacional Linux, para ajudar aqueles que não estejam confortáveis com este sistema operacional.



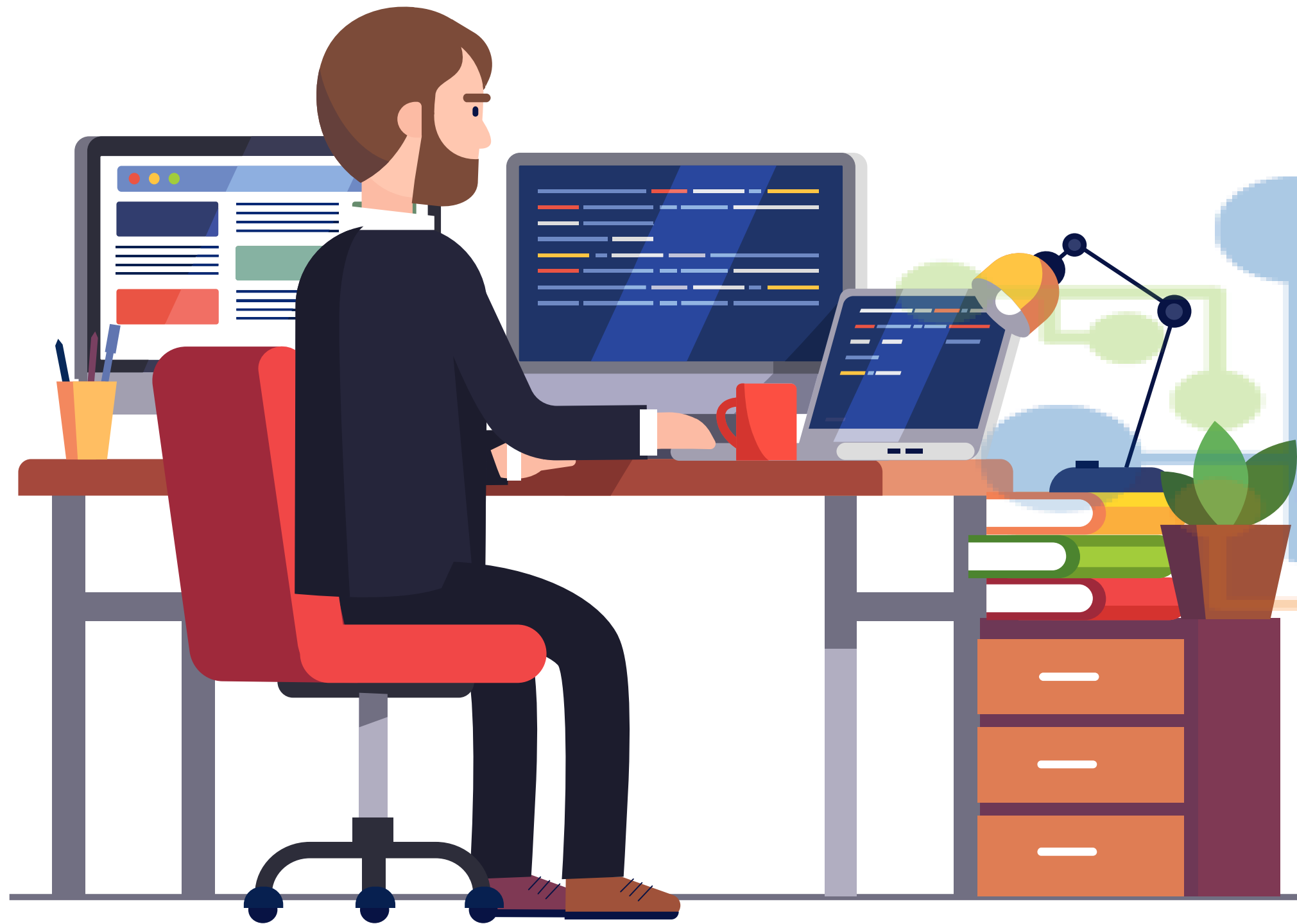
Big Data  
Fundamentos 2.0 e  
Introdução à Ciência  
de Dados 2.0



Sistema Operacional  
Linux

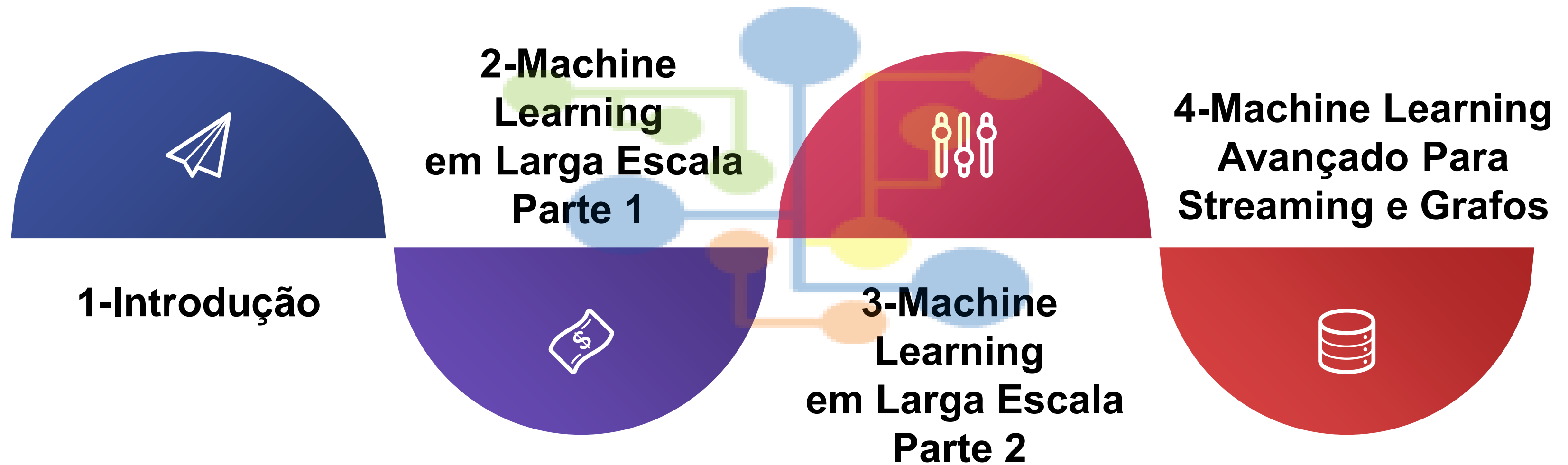


Ambiente em Nuvem  
AWS



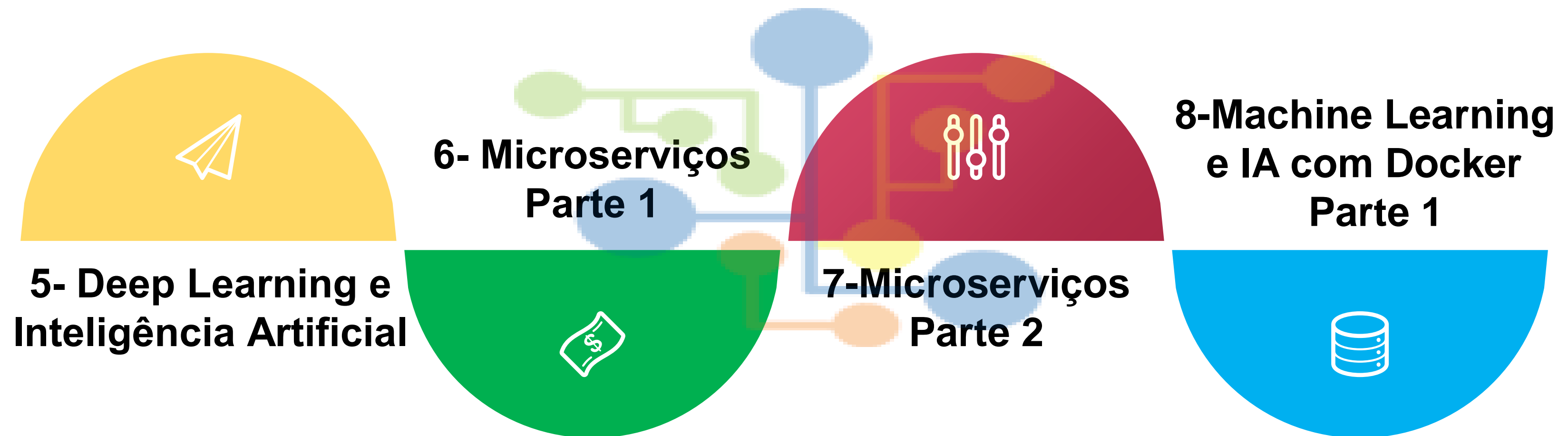


# Módulos





# Módulos







# Módulos





# O que Esperar Deste Curso?

- Teoria e Prática
- Dinamismo
- Leitura
- Muito Conteúdo





# Materiais de Aprendizagem



## Aulas em Vídeo

Exposição sobre o conteúdo



## Laboratórios

Cenários e  
Troubleshooting



## Pesquisa Adicional

Bibliografia, referências e links úteis ao final de cada capítulo



## Quizzes e Exercícios

Quizzes e exercícios para testar seu conhecimento





# O que Esperamos de Você?

## Divirta-se

Comunique-se, aprenda e divirta-se em nossa Comunidade.

## Bibliografia

Leia a bibliografia adicional, acesse os links úteis e realize os quizzes ao final dos capítulos.



**2 a 4 horas**

de dedicação por semana.

## Leitura do Material

E-books e material complementar!

## Interação

Utilize nossas Apps e interaja na rede com outros alunos, no fórum exclusivo e na timeline da Comunidade em nosso site.

# Avaliação Final



3 Tentativas

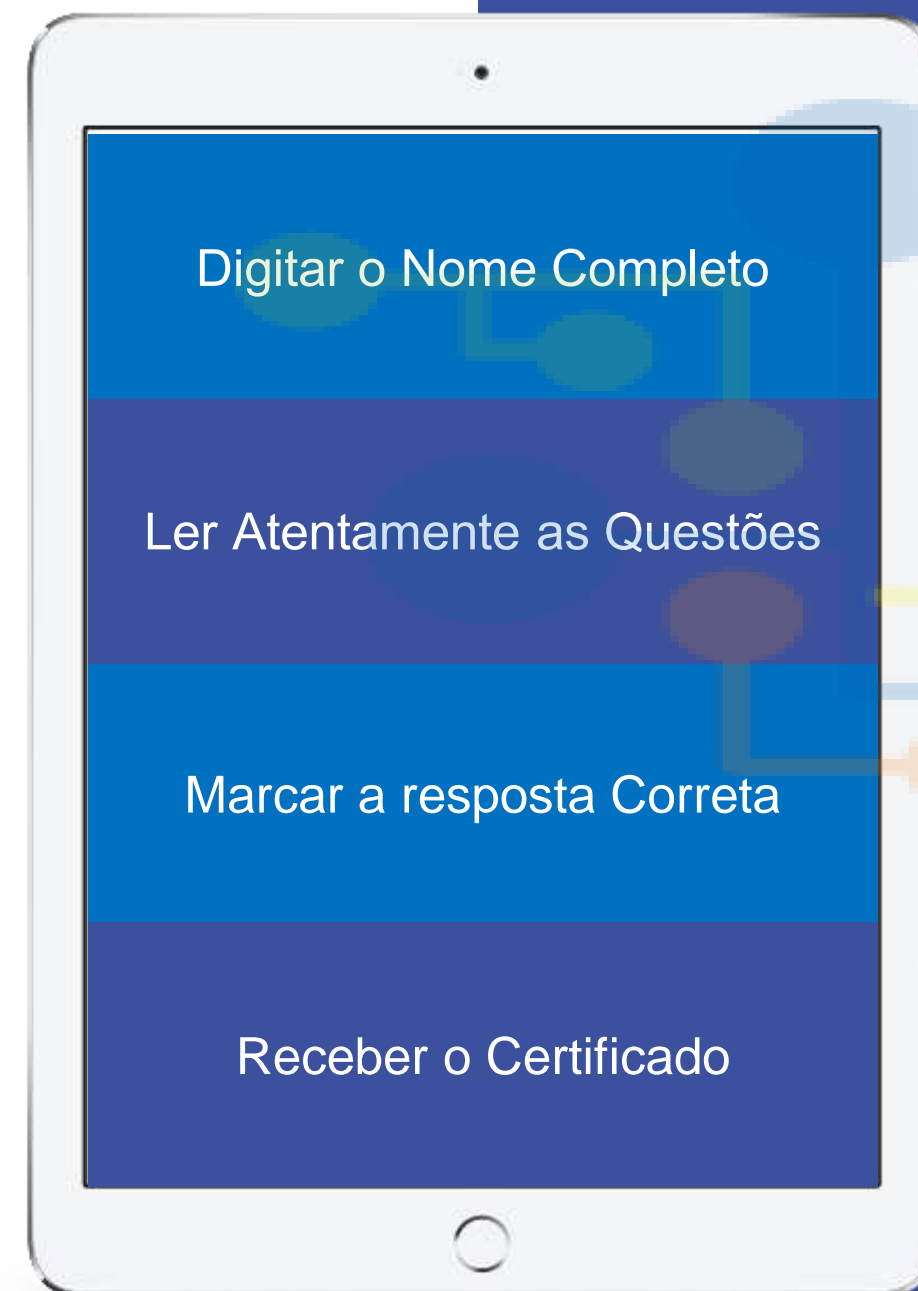


50 Questões



70 %

Aproveitamento





# Bibliografia

## Bibliografia Sugerida







Data Science  
Academy

Data Science Academy eng.davidborges@gmail.com 59532d8f5e4cdead748b456a

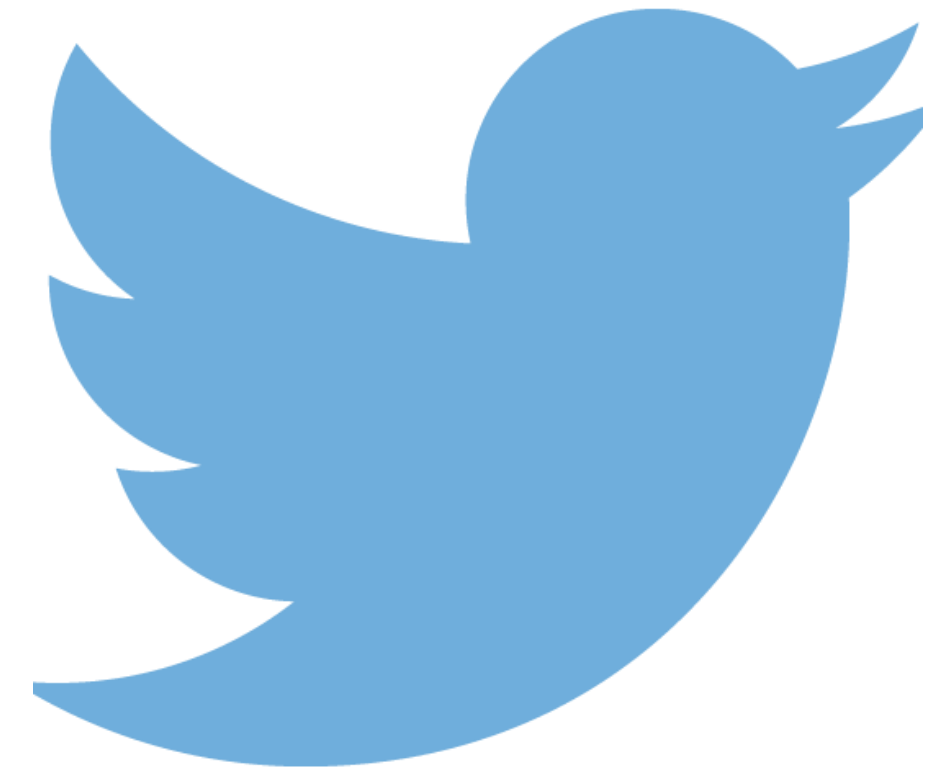
# APPs Gratuitas





# Compartilhe Seu Certificado De Conclusão

LinkedIn







Data Science  
Academy

Data Science Academy eng.davidborges@gmail.com 59532d8f5e4cdead748b456a

# Inteligência Artificial

x

# Machine Learning

x

# Deep Learning

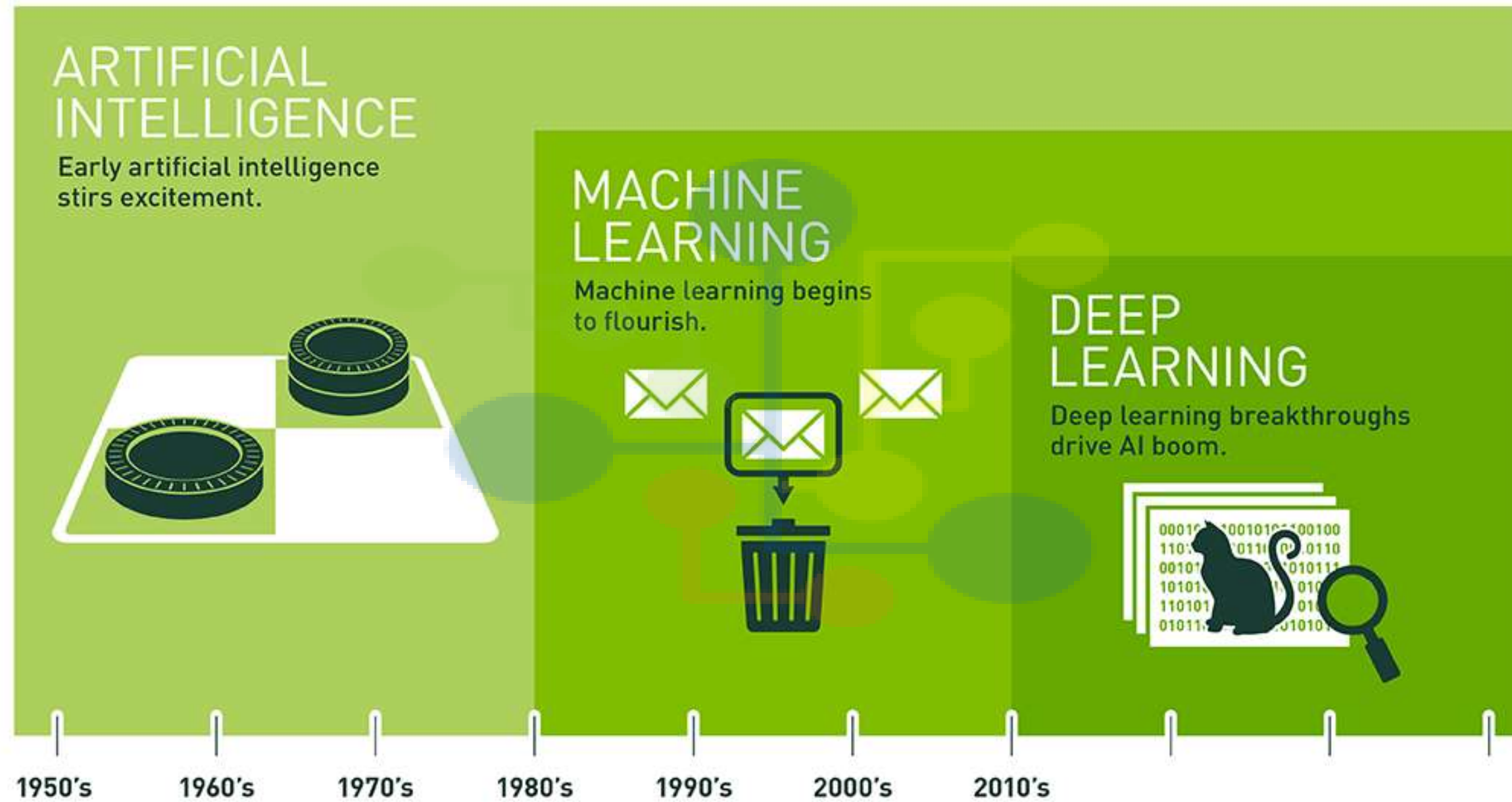


Data Science Academy





# Inteligência Artificial x Machine Learning x Deep Learning



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.





Data Science  
Academy

Data Science Academy eng.davidborges@gmail.com 59532d8f5e4cdead748b456a

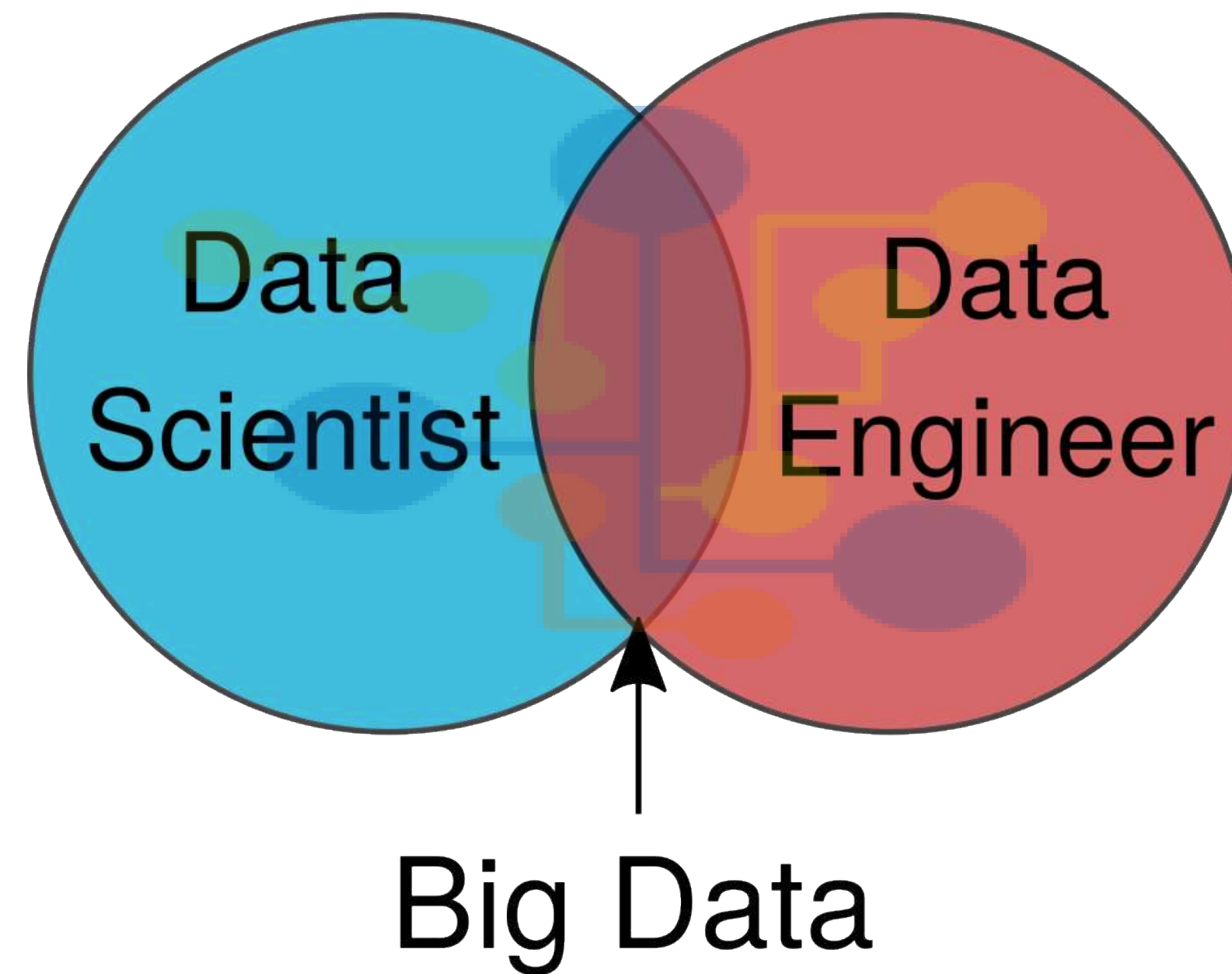
# Qual o Papel do Engenheiro de Dados no Processo de Machine Learning?



Data Science Academy



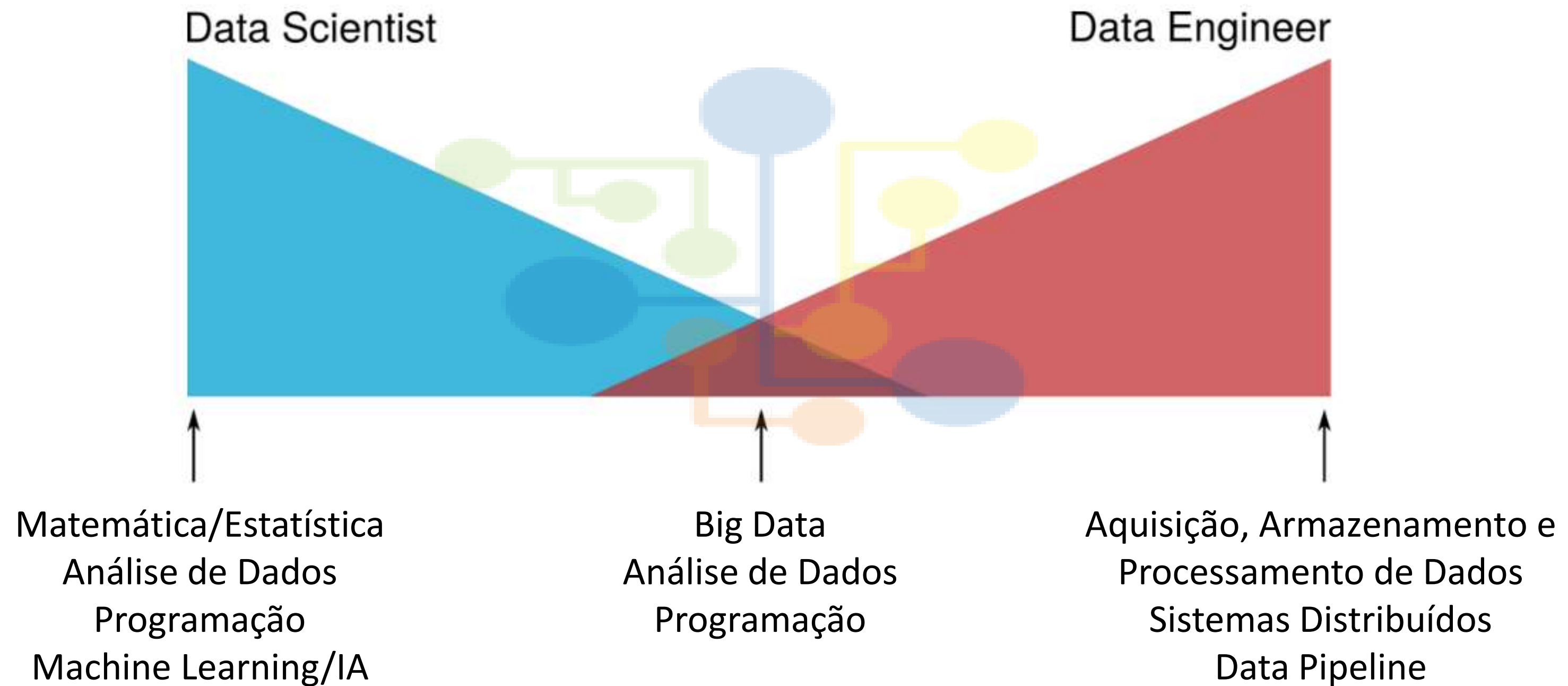
# Qual o Papel do Engenheiro de Dados no Processo de Machine Learning?







# Qual o Papel do Engenheiro de Dados no Processo de Machine Learning?





Data Science  
Academy

Data Science Academy eng.davidborges@gmail.com 59532d8f5e4cdead748b456a

# Processamento Paralelo em Ambientes Distribuídos



Data Science Academy



# Processamento Paralelo em Ambientes Distribuídos

Um sistema de processamento distribuído ou paralelo é um sistema que interliga vários nós de processamento (computadores individuais, não necessariamente homogêneos) de maneira que um processo de grande consumo seja executado no nó "mais disponível", ou mesmo subdividido por vários nós conseguindo-se, portanto, ganhos óbvios nestas soluções: uma tarefa qualquer, se divisível em várias subtarefas pode ser realizada em paralelo.

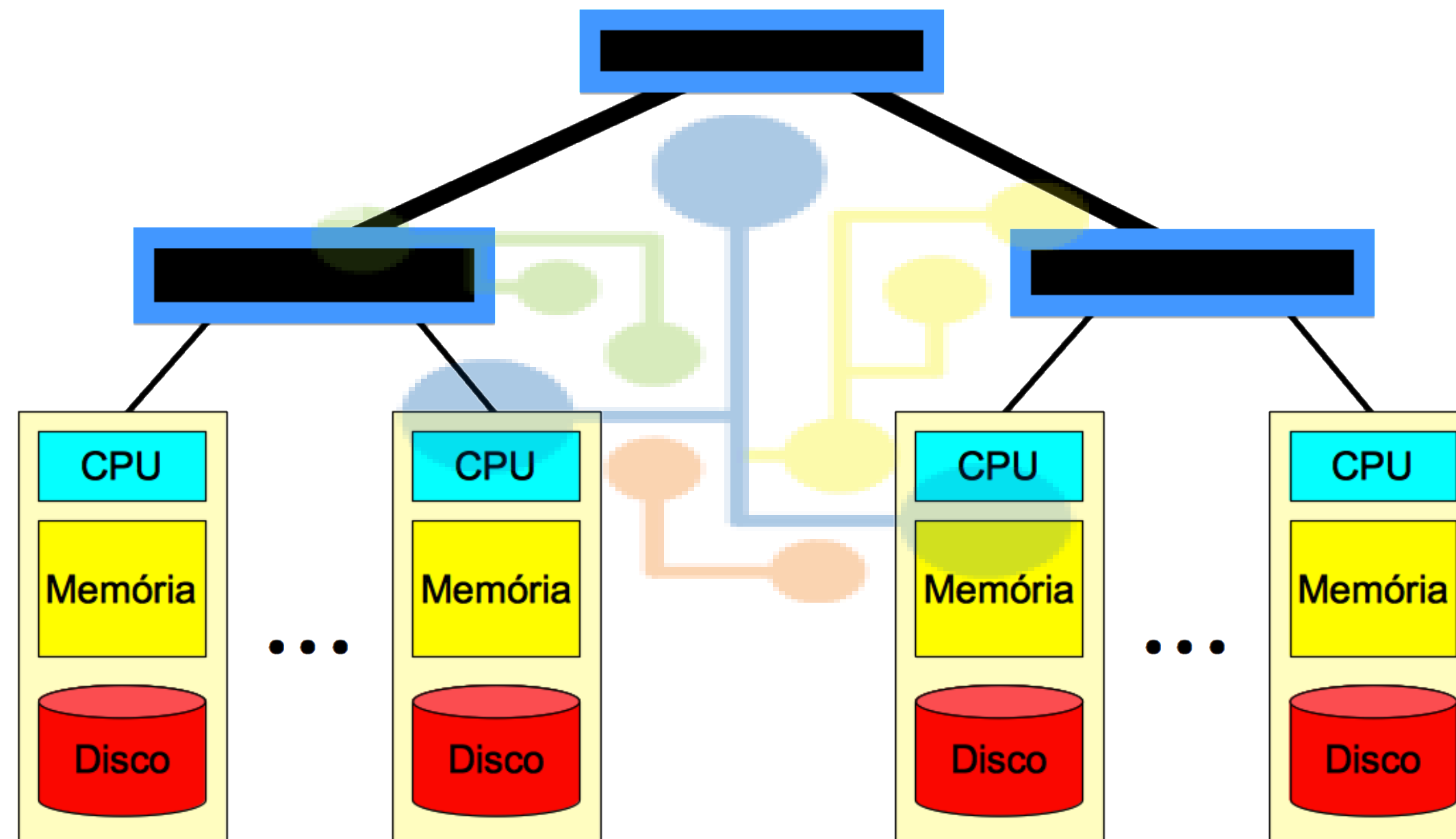




# Processamento Paralelo em Ambientes Distribuídos

A nomenclatura geralmente utilizada neste contexto é:  
HPC (High Performance Computing) e/ou DPC (Distributed/Parallel Computing).

# Processamento Paralelo em Ambientes Distribuídos



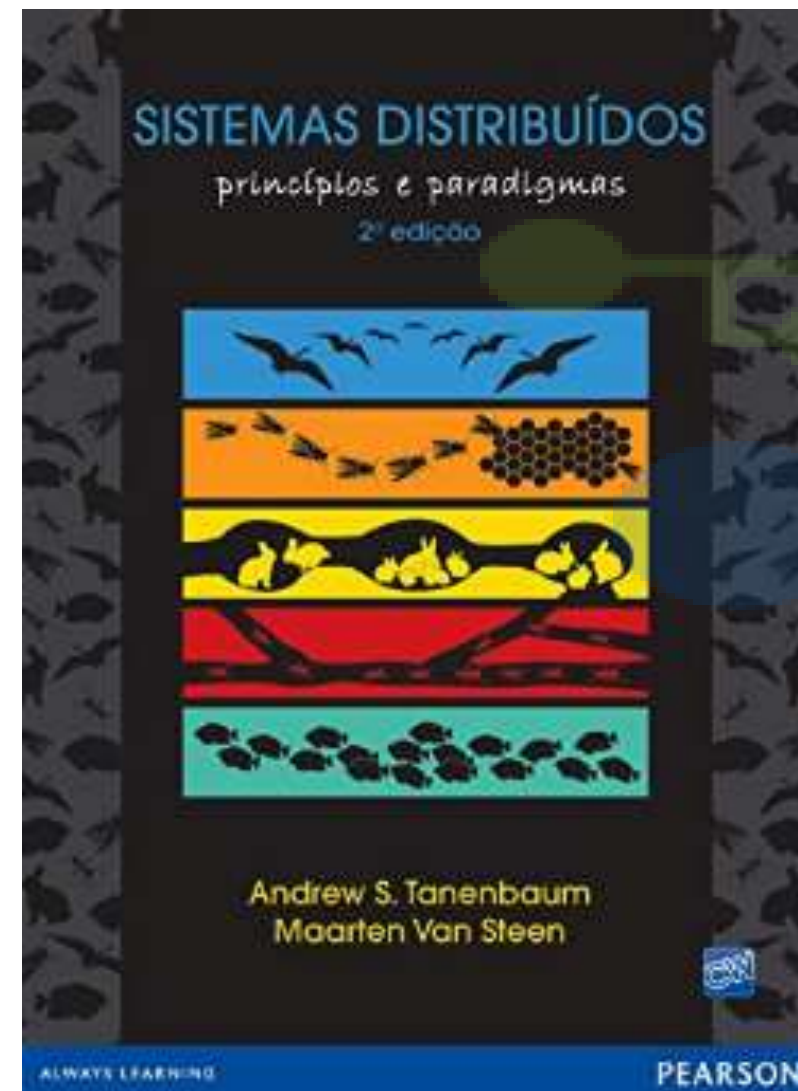


# Processamento Paralelo em Ambientes Distribuídos

Um sistema distribuído segundo a definição de Andrew Tanenbaum é uma:  
"coleção de computadores independentes entre si que se apresenta ao usuário como um sistema único e coerente".



# Processamento Paralelo em Ambientes Distribuídos







Data Science  
Academy

Data Science Academy [eng.davidborges@gmail.com](mailto:eng.davidborges@gmail.com) 59532d8f5e4cdead748b456a

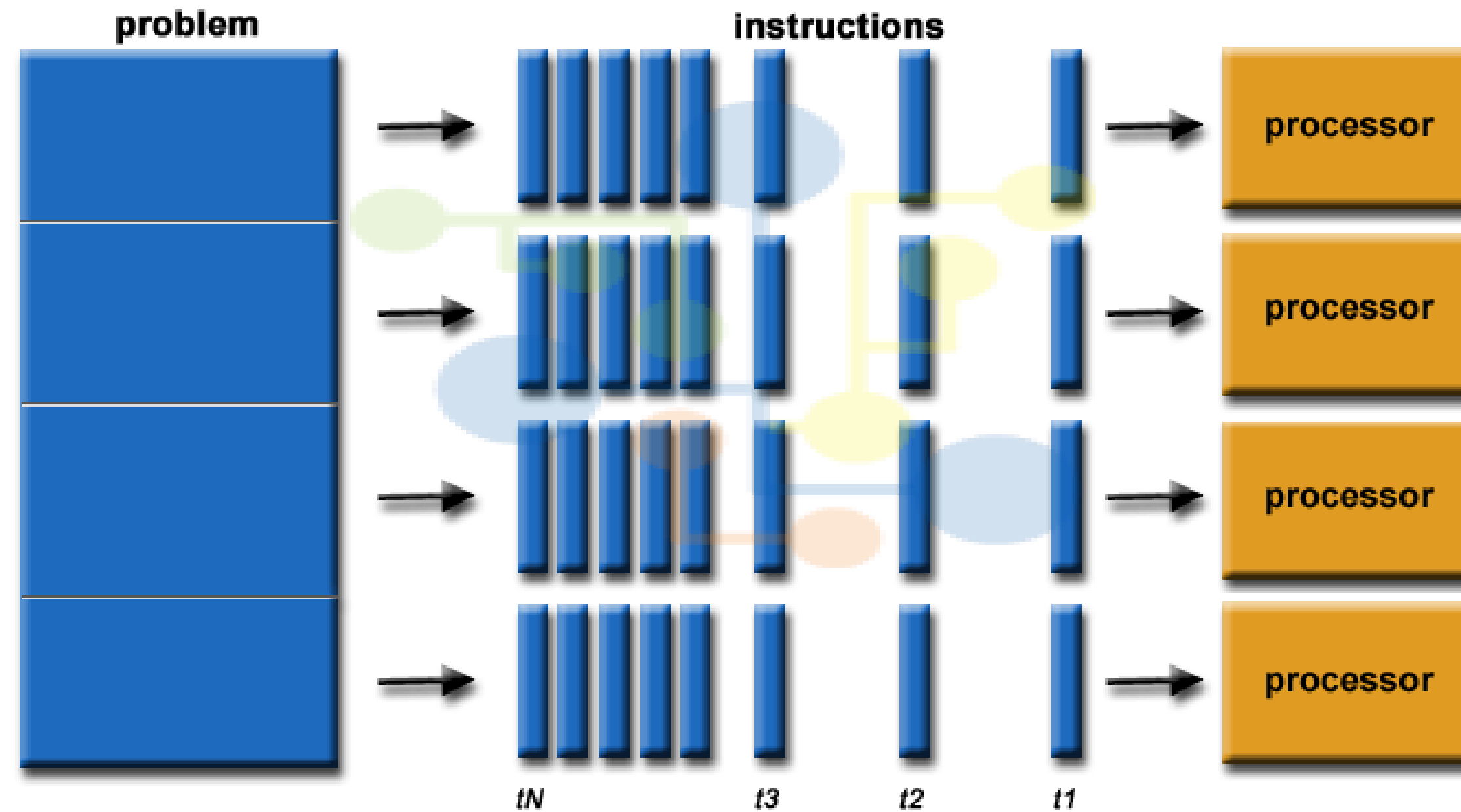
# Diferenças Entre CPU x GPU



Data Science Academy

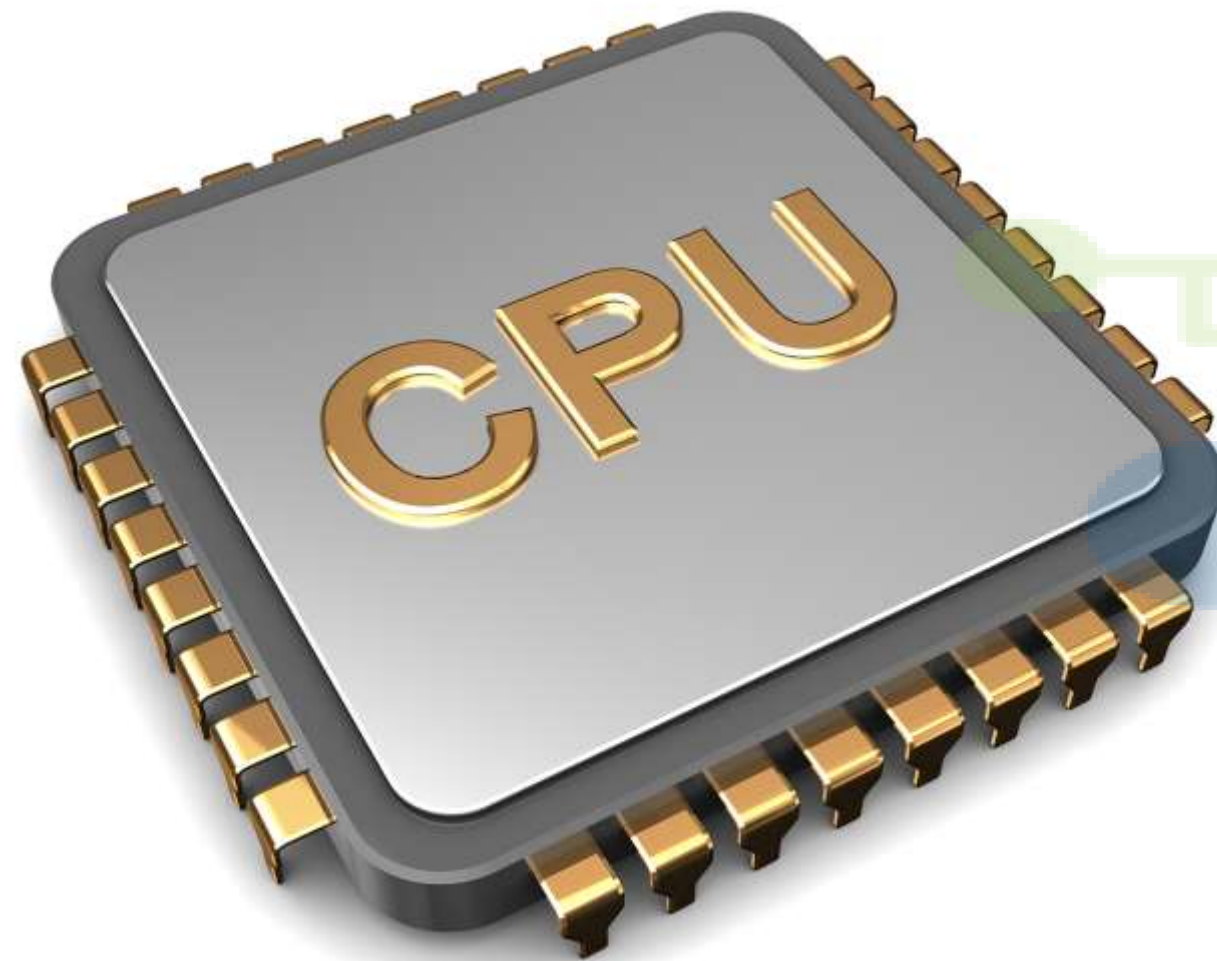


# Diferenças Entre CPU x GPU





# Diferenças Entre CPU x GPU







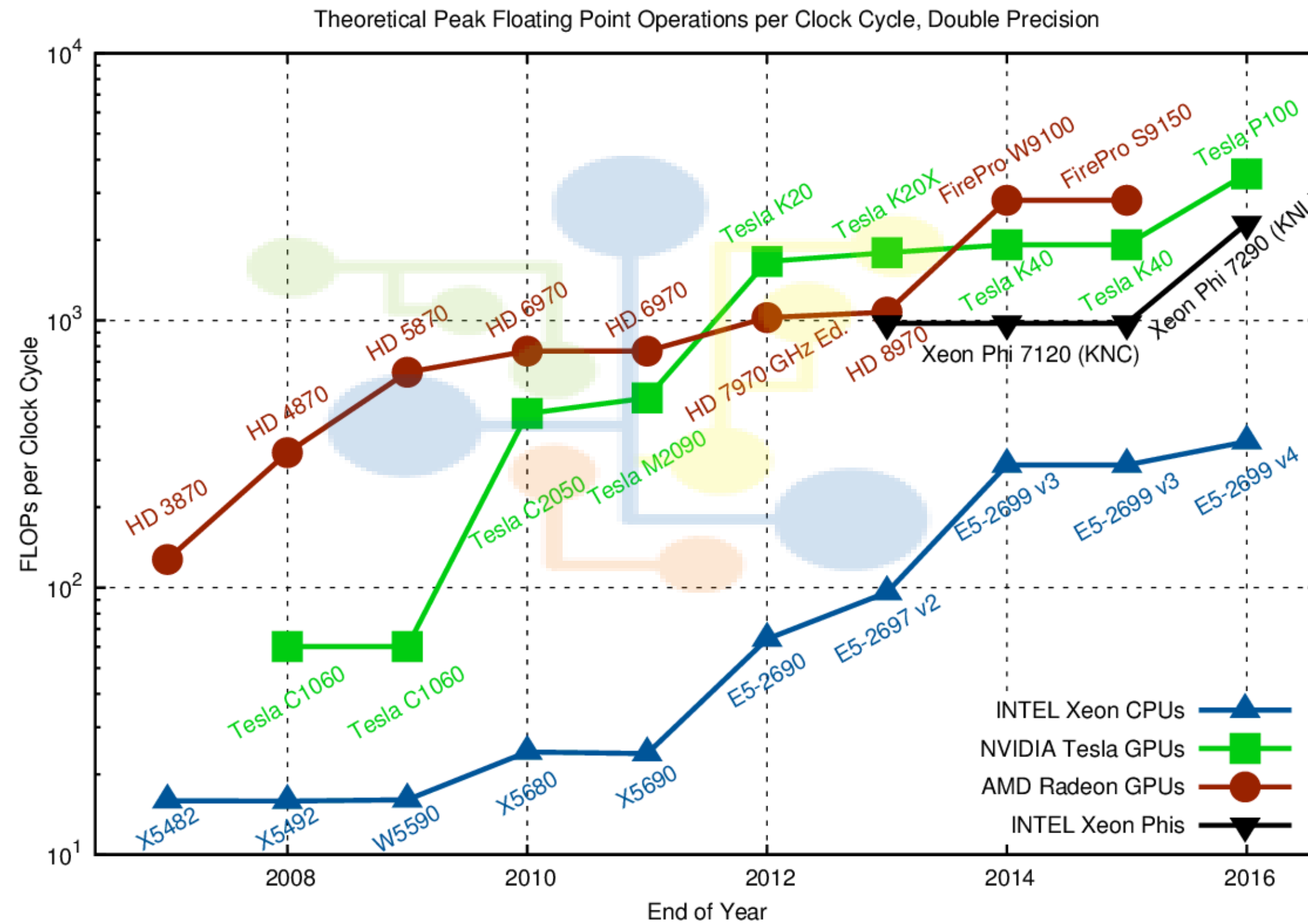
## Diferenças Entre CPU x GPU



# NVIDIA



# Diferenças Entre CPU x GPU







Data Science  
Academy

Data Science Academy [eng.davidborges@gmail.com](mailto:eng.davidborges@gmail.com) 59532d8f5e4cdead748b456a

# O que é Programação Paralela?



Data Science Academy





# O que é Programação Paralela?

**Programação  
Paralela**

**Processamento  
Paralelo**

**Ambientes  
Distribuídos**

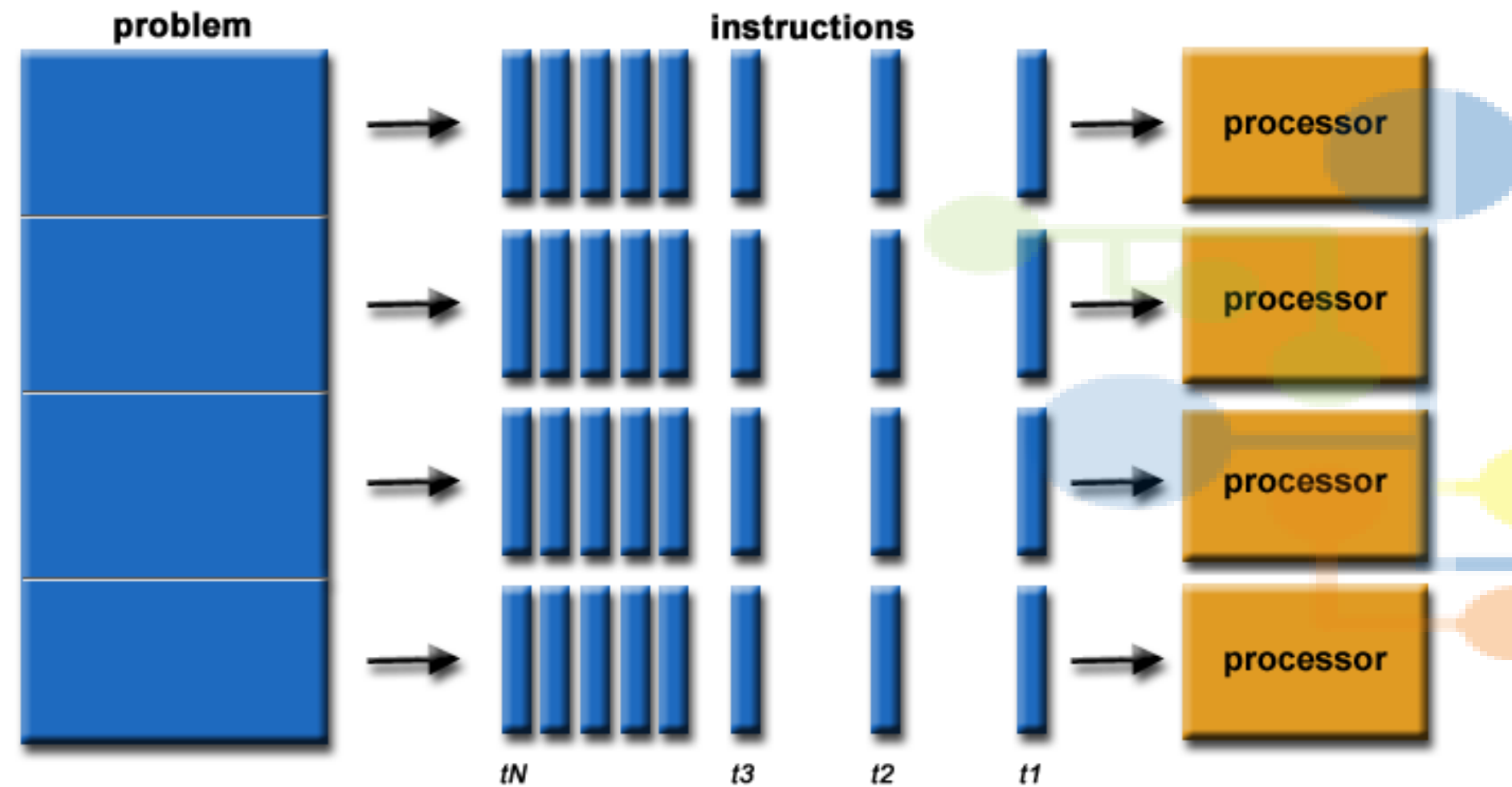


# O que é Programação Paralela?

**Concorrência** é quando um servidor atende a vários clientes escalando um determinado tempo para atender cada um, já o **Paralelismo** é quando vários servidores atendem vários clientes ao mesmo tempo, reduzindo o tempo de resposta para os clientes.



# O que é Programação Paralela?




Escrever programas que utilizem paralelismo é um pouco diferente de escrever programas para execução sequencial, como estamos mais acostumados a fazer. Temos que escolher escolher que partes do programa utilizarão código que serão executados em paralelo.





# O que é Programação Paralela?



Sempre que dois processamentos forem independentes, por definição, se pode usar paralelização.

# O que é Programação Paralela?

**Um exemplo de problema facilmente paralelizável é:**

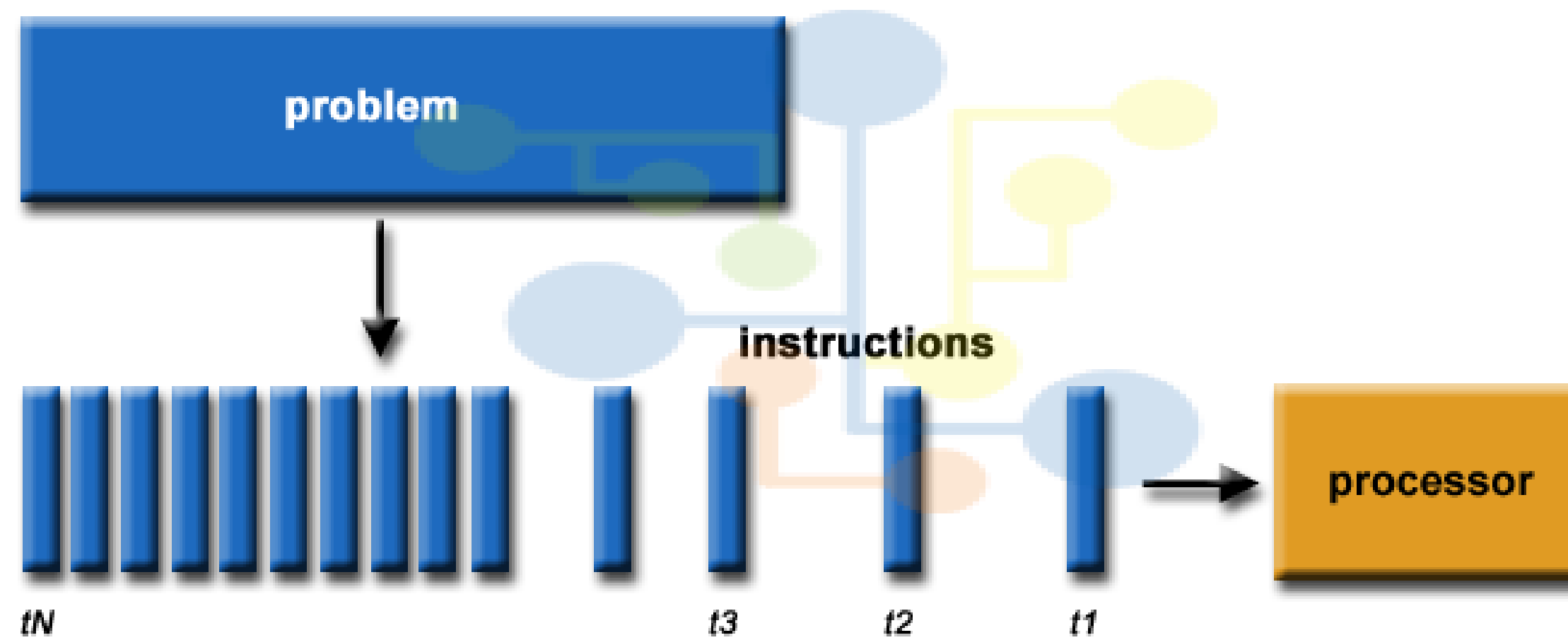
"divida o vetor em N partes e mande cada processo realizar a tarefa em uma das partes"

**Um exemplo de problema dificilmente paralelizável é:**

"faça com que cada elemento desse vetor enorme tenha seu valor somado a todos os elementos posteriores a ele"



# O que é Programação Paralela?







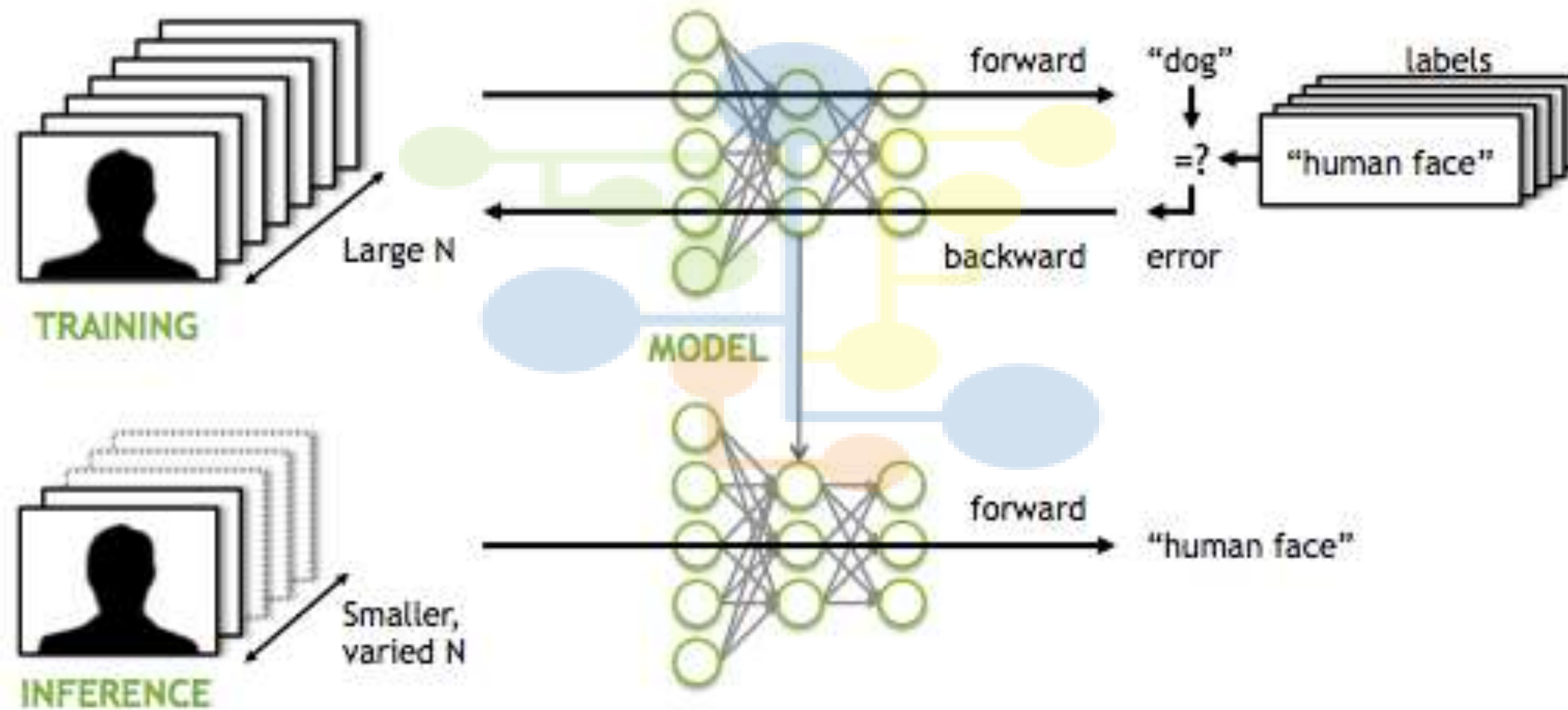
# O que é Programação Paralela?

E quando usar Programação Paralela?





# O que é Programação Paralela?

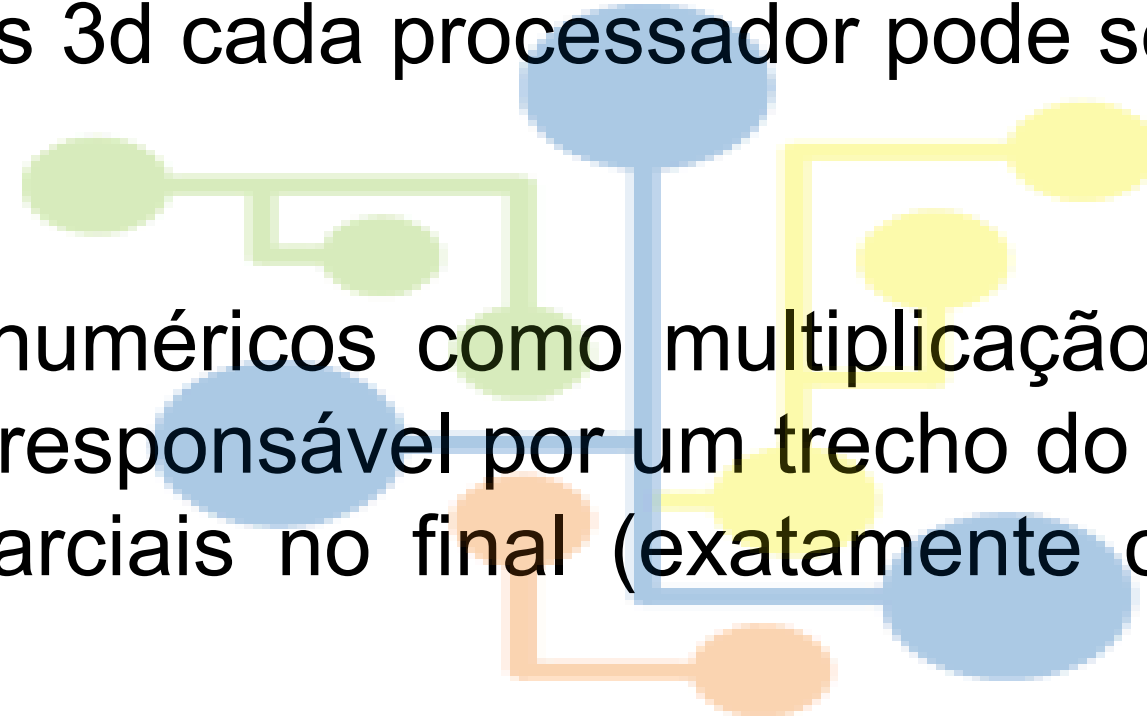




# O que é Programação Paralela?

Computações "obviamente paralelizáveis":

- Para renderizar imagens 3d cada processador pode ser responsável por um pedaço diferente da tela.
- Em alguns algoritmos numéricos como multiplicação de vetores e matrizes, cada processador pode ficar responsável por um trecho do vetor ou matriz separado e dá pra juntar as contas parciais no final (exatamente o que ocorre em modelos de Deep Learning).
- Para compilar um programa com vários módulos, é possível usar paralelismo para compilar mais de um módulo ao mesmo tempo.







Data Science  
Academy

Data Science Academy [eng.davidborges@gmail.com](mailto:eng.davidborges@gmail.com) 59532d8f5e4cdead748b456a

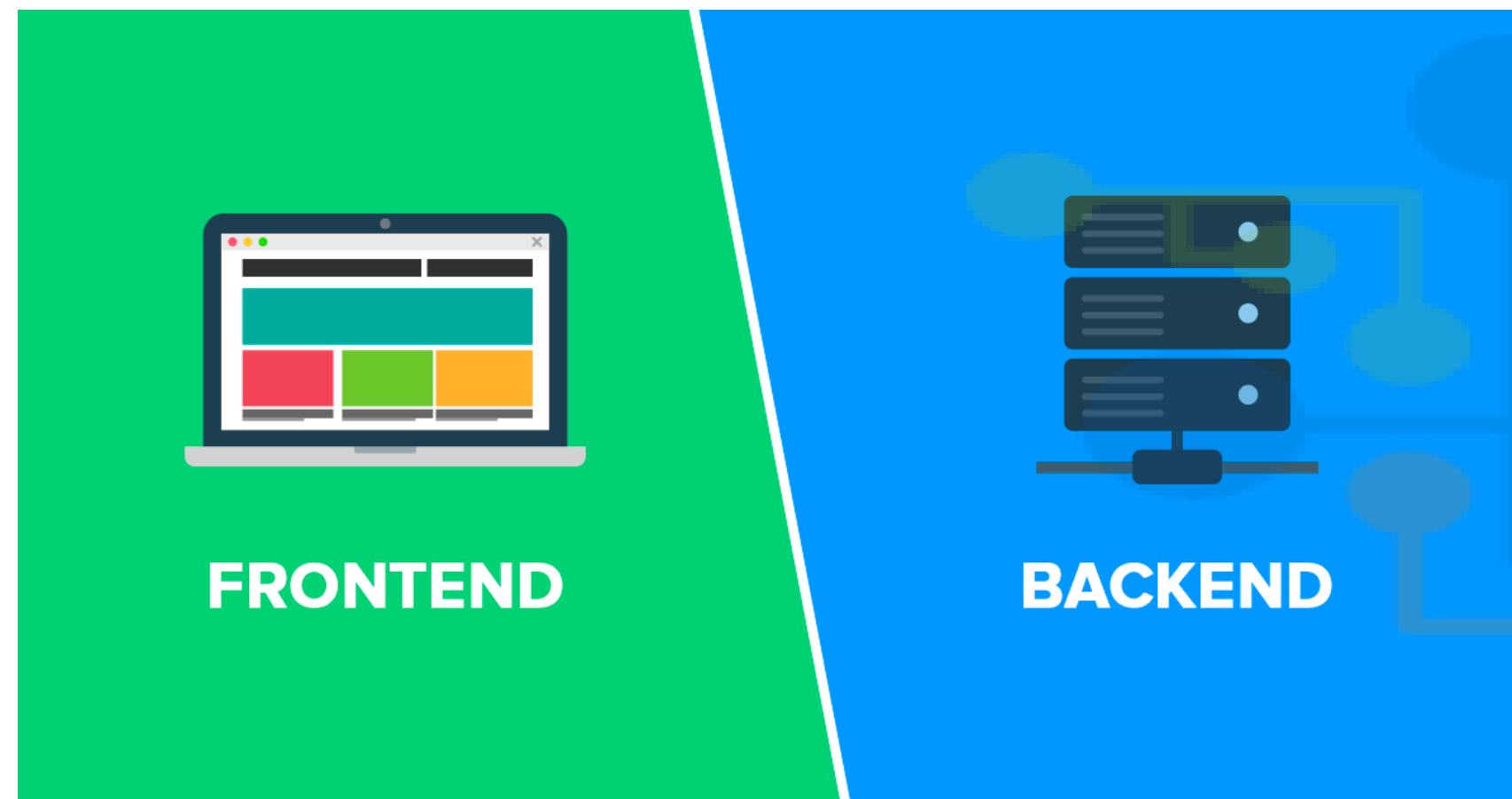
# O que são Microservices?



Data Science Academy



# O que são Microservices?

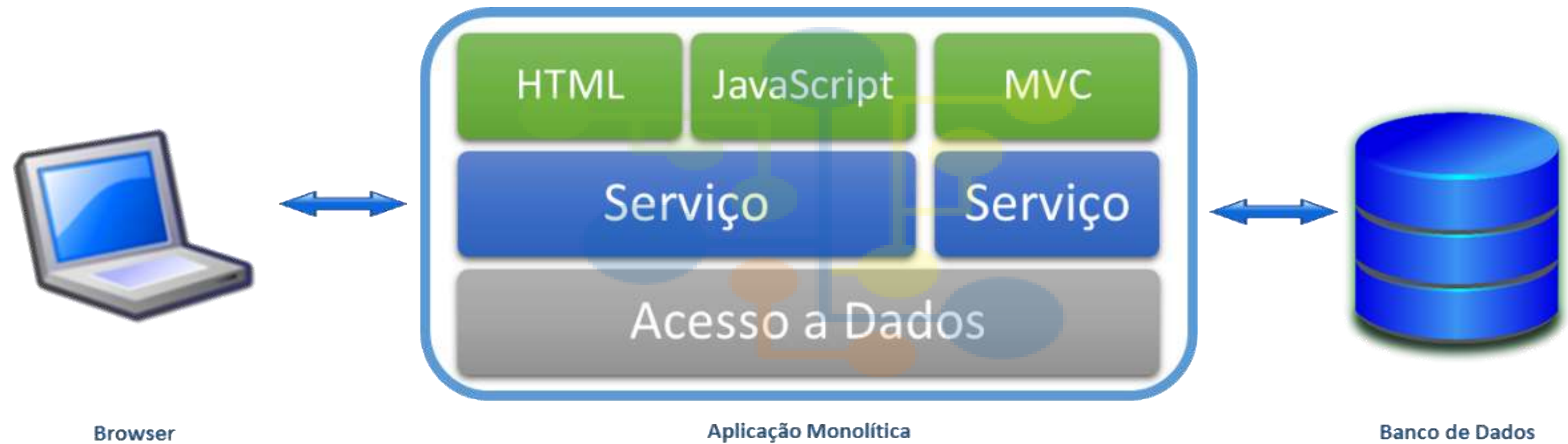


Antes de compreender o que são Microservices, vamos compreender como funciona a arquitetura padrão de aplicações (softwares).





# O que são Microservices?







# O que são Microservices?





# O que são Microservices?

## Vantagens de uma aplicação monolítica

- Simplicidade da arquitetura: Não existem muitas camadas com o que se preocupar.
- Agregação de tecnologia: Toda a aplicação é desenvolvida em uma mesma tecnologia, facilitando a coesão da equipe.
- Fluxo de publicação simples: Alterou? Compilou? É só publicar.
- Rápido desenvolvimento: Por ser uma arquitetura mais simples, o seu desenvolvimento tende a ser muito mais rápido.



# O que são Microservices?

## Desvantagens de uma aplicação monolítica

- Agregação de tecnologia.
- Único ponto de falha: problema no sistema de newsletters? Não conseguimos fazer o pagamento dos funcionários porque o sistema de folha não funciona.
- Baixa escalabilidade: temos que copiar TODA a stack para escalar horizontalmente.
- Base de código gigante: quanto maior o sistema, maior é a base de código, já que está tudo no mesmo lugar.
- Desperdício de esforço: imagine que você tenha que mudar o texto de uma das telas de uma aplicação monolítica de 50.000 linhas, quando você for publicar isso, ela vai ter que ser totalmente recompilada, tudo isso por causa de um texto, será que valeu a pena?



# O que são Microservices?

O Microservices, também conhecido como arquitetura de microserviços, é um estilo arquitetônico que estrutura uma solução como uma coleção de serviços ligeiramente acoplados, que implementam capacidades empresariais.

# O que são Microservices?

"...Uma abordagem que desenvolve um aplicativo único como uma suíte de pequenos serviços..." (Martin Fowler, 2014)

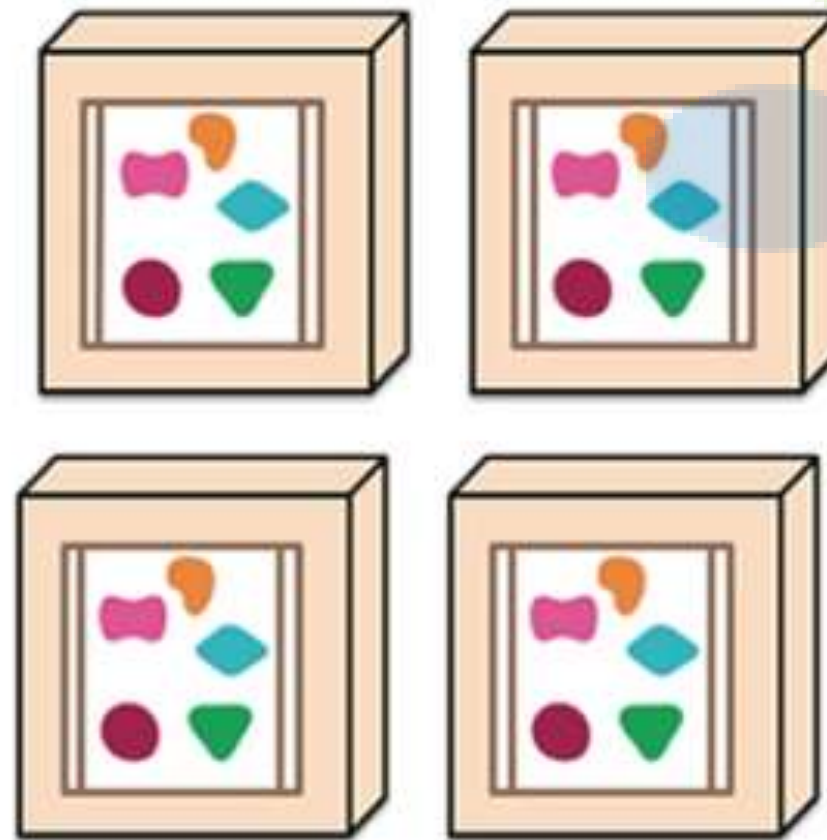


# O que são Microservices?

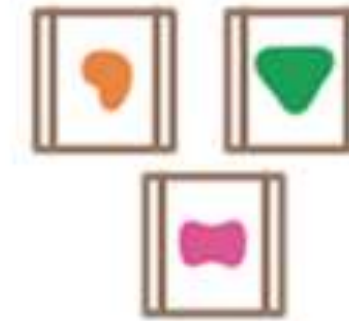
*Um aplicativo monolítico tem todas as suas funcionalidades em um único processo...*



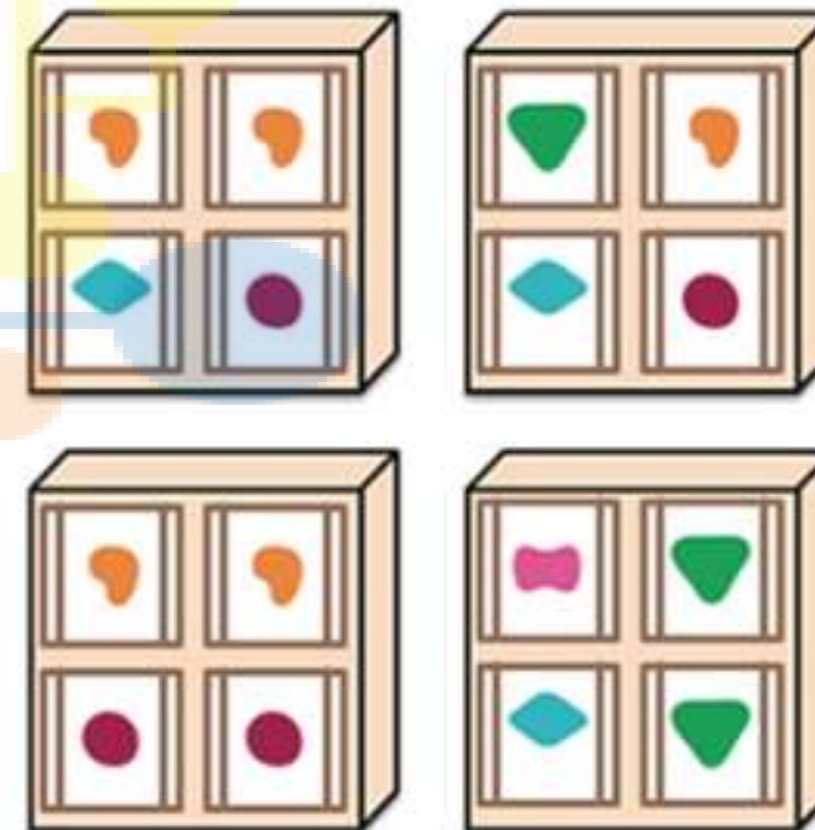
*...e escala replicando o monolito em servidores múltiplos*



*A arquitetura de micro-serviços coloca cada elemento de funcionalidade em um serviço separado...*



*...e escala distribuindo os serviços entre os servidores, replicando por demanda.*



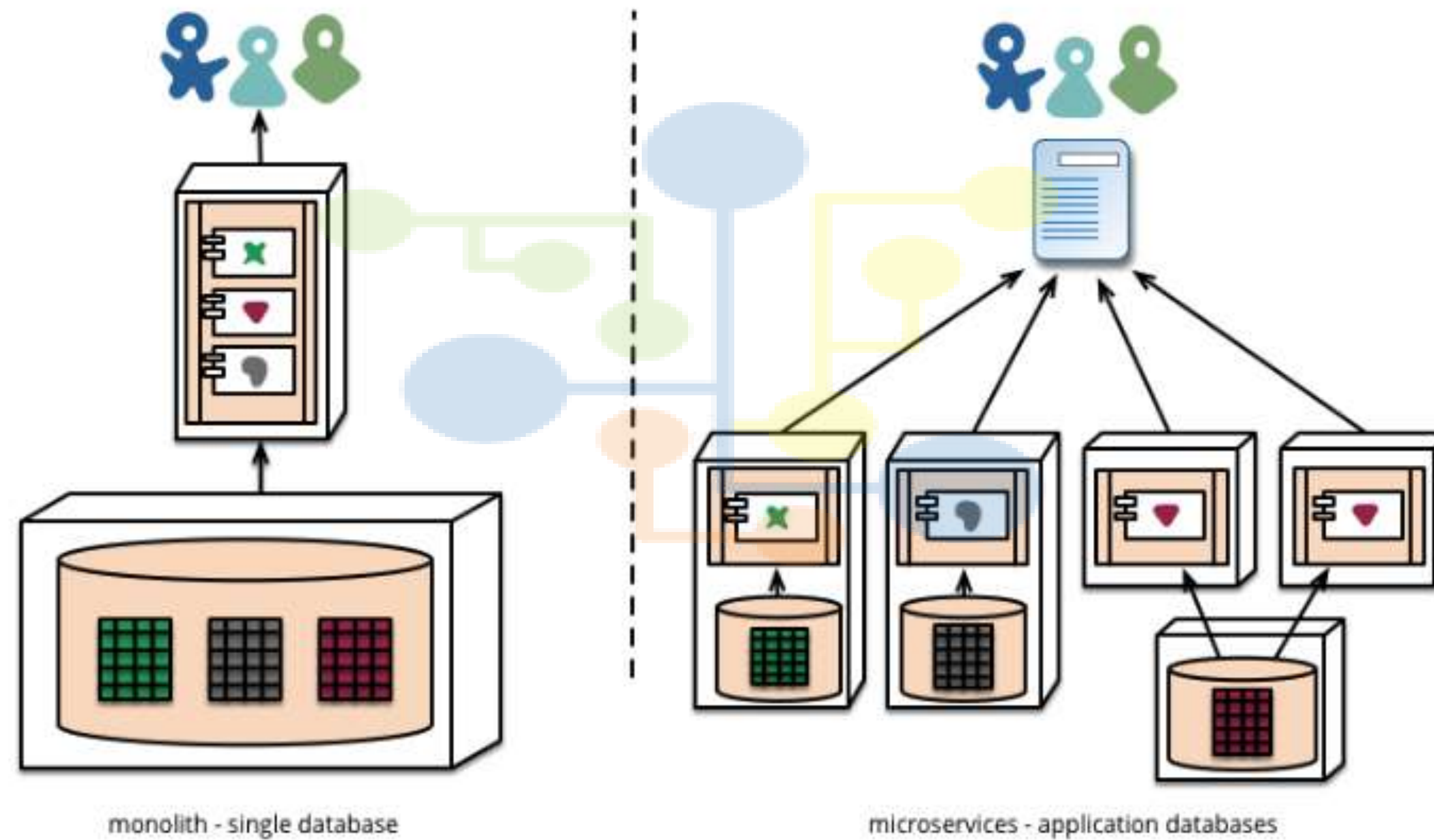


# O que são Microservices?

Em geral, um microserviço é um sistema simples (geralmente uma API) que se comunica através de mecanismos leves (como o HTTP). Estes pequenos sistemas devem ser totalmente autônomos, ou seja, devem possuir sistemas de deploy totalmente independentes e totalmente automáticos.



# O que são Microservices?

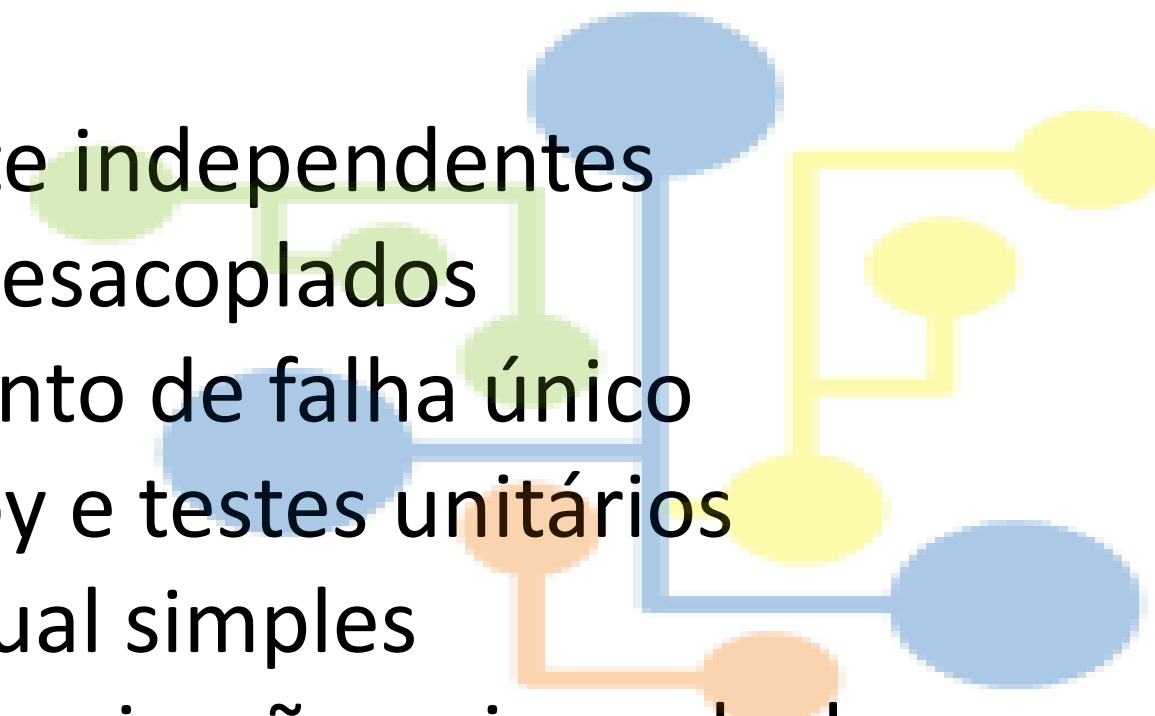




# O que são Microservices?

## Vantagens dos Microservices

- Sistemas totalmente independentes
- Serviços coesos e desacoplados
- Ausência de um ponto de falha único
- Facilidade de deploy e testes unitários
- Arquitetura individual simples
- Mecanismo de comunicação universal e leve
- Não se prende a uma tecnologia específica







# O que são Microservices?

## Desvantagens

- A arquitetura geral pode se tornar complexa se não for bem documentada.
- Se não for bem planejado e bem executado, pode se transformar em uma grande bagunça.
- A ausência de uma documentação que descreva o escopo global pode onerar a visualização geral do sistema.
- Em um início de projeto, os problemas resolvidos por esta arquitetura são inexistentes. Portanto um dos grandes desafios é saber em qual parte do ciclo do projeto ela deve ser implementada, isto demanda mais tempo de desenvolvimento.
- O desenvolvimento de recursos novos que abrangem mais de um serviço devem ser cuidadosamente orquestrados em todas as equipes.



# O que são Microservices?

## Infraestrutura

Kubernetes: Sistema de orquestração de containers Docker. Ideal para a orquestração de microserviços pois é baseado totalmente em API's e ele próprio é dividido em diversos hosts (se tornando um microserviço em si), muito fácil de instalar e utilizar. Agnóstico de infraestrutura, pode ser executado em diversos provedores.

AWS ECS: Serviço de containers hospedado pela AWS (muito similar ao kubernetes, em sua própria forma).

Apache Mesos + Marathon + ZooKeeper: Abstrai as configurações de diversos computadores e transforma em um único "super super computador".





Data Science  
Academy

Data Science Academy [eng.davidborges@gmail.com](mailto:eng.davidborges@gmail.com) 59532d8f5e4cdead748b456a

# O que São REST APIs?



Data Science Academy





# O que São REST APIs?





# O que São REST APIs?

É um estilo de arquitetura para transferência de informação, que realiza operações no padrão HTTP.

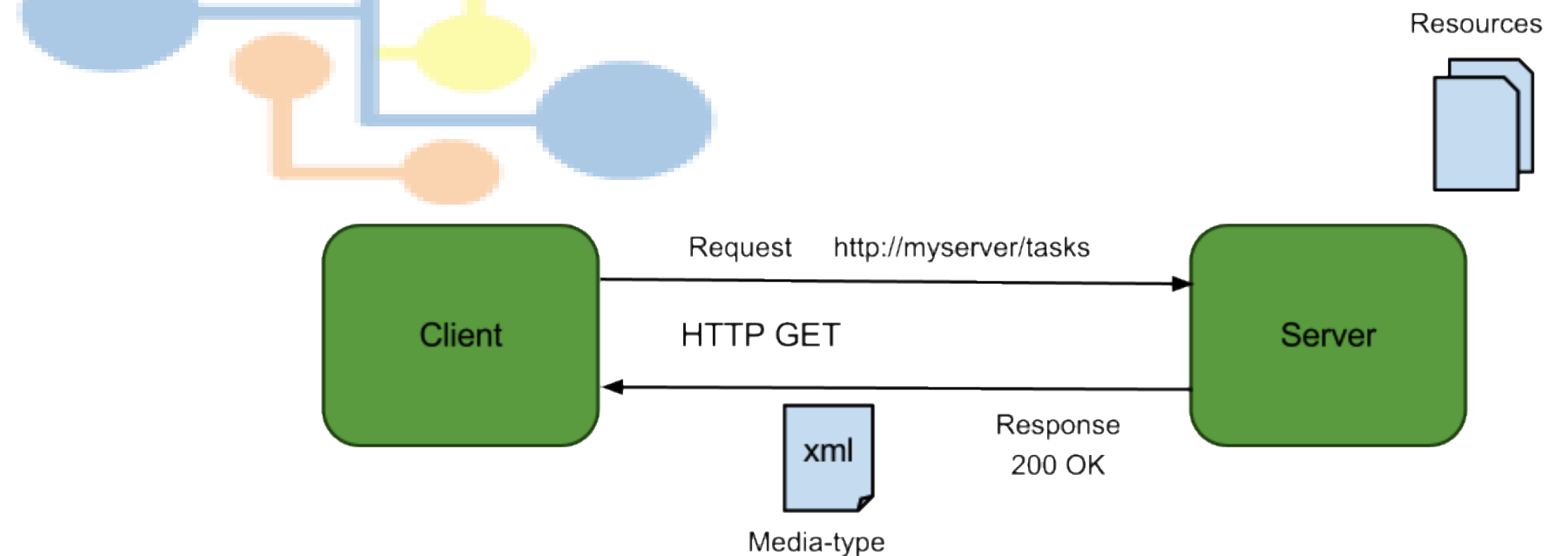
Pode tratar diferentes tipos de dados, sendo XML e JSON os mais comuns.





## O que São REST APIs?

REST é um termo definido por Roy Fielding em sua tese de mestrado no qual ele descreve sobre um estilo de arquitetura de software sobre um sistema operado em rede. REST é um acrônimo para "Transferência de Estado Representacional" (Representational State Transfer).

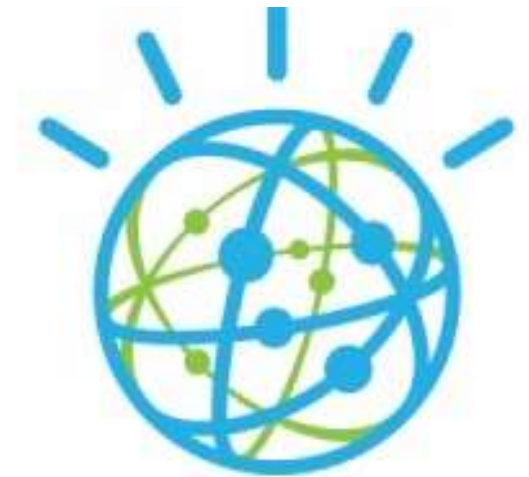






## O que São REST APIs?

O IBM Watson oferece suas funcionalidades de forma que possam ser integradas em outras aplicações (as aplicações da sua empresa por exemplo) através de APIs com a arquitetura REST e com frameworks leves e de alta capacidade.



**IBM Watson**



Data Science  
Academy

Data Science Academy eng.davidborges@gmail.com 59532d8f5e4cdead748b456a

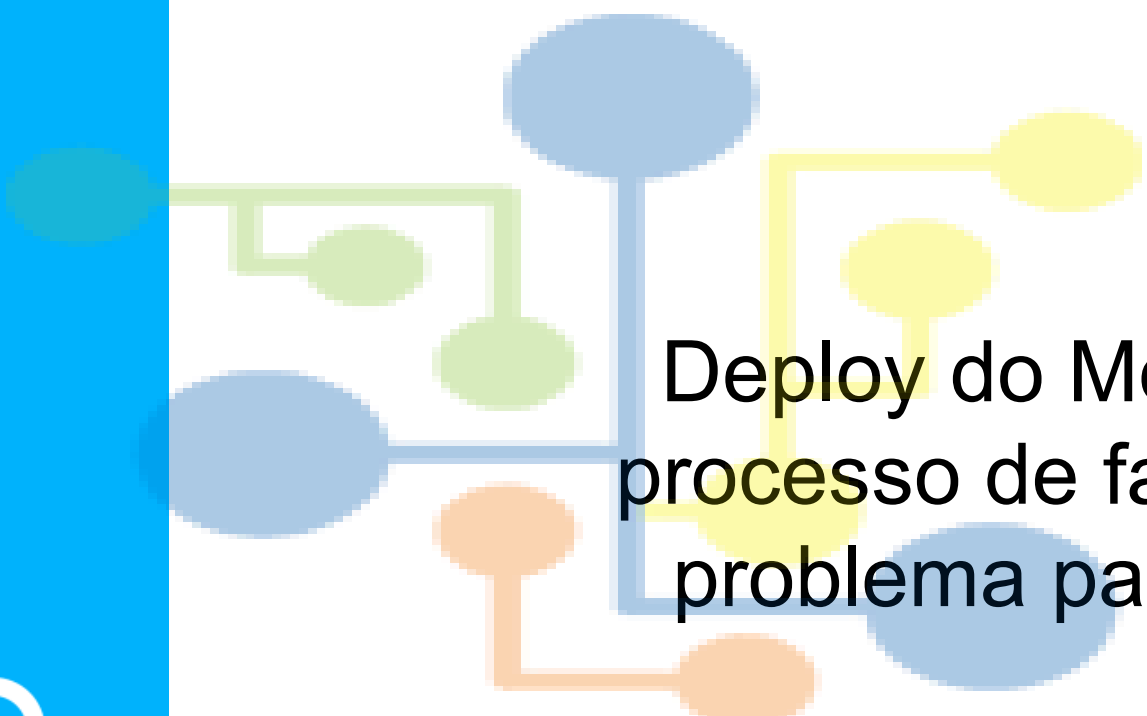
# Como Publicar Modelos de Machine Learning em Produção



Data Science Academy



# Como Publicar Modelos de Machine Learning em Produção

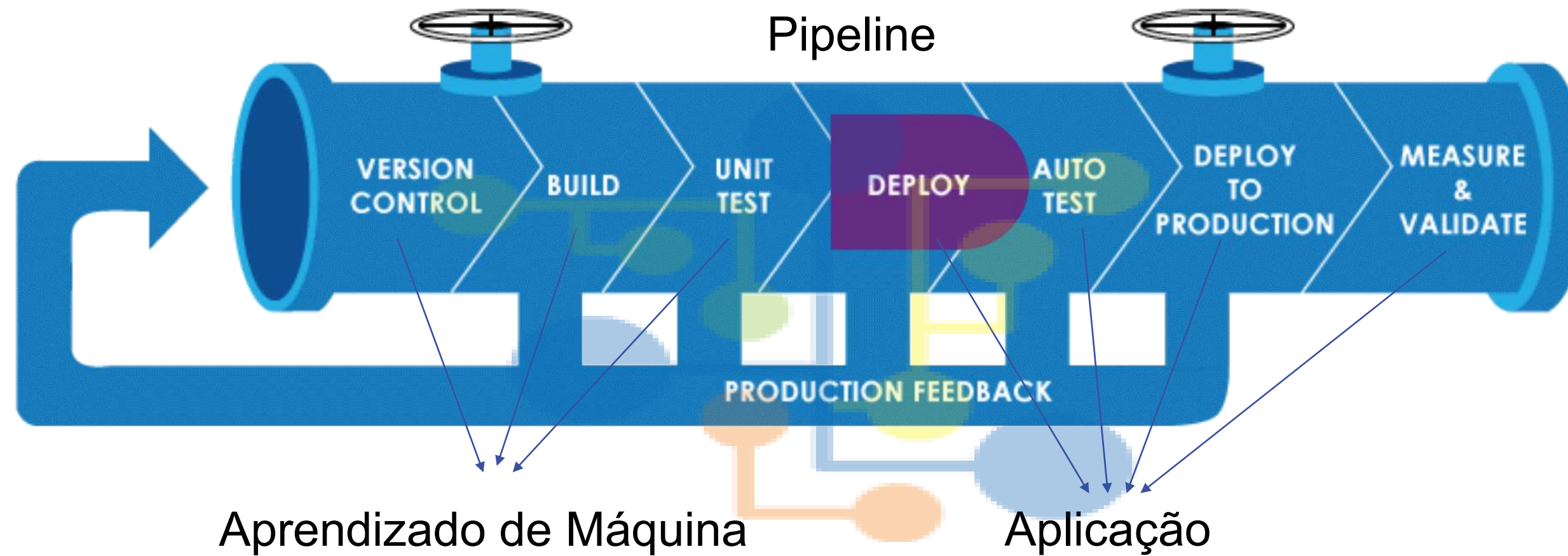
A decorative network diagram with several nodes of different colors (blue, green, yellow, orange) connected by lines, positioned behind the text.

Deploy do Modelo em Produção é o processo de fazer o modelo resolver o problema para o qual ele foi criado.





# Como Publicar Modelos de Machine Learning em Produção



Cientista de Dados  
desenvolve os modelos

Desenvolvedor cria aplicações  
Engenheiro de Dados faz o deploy de todo o Pipeline



# Como Publicar Modelos de Machine Learning em Produção

**Aprendizado de Máquina é um meio e não um fim.** Para fazer o Deploy do modelo em produção, você precisa criar uma aplicação que faça a leitura do resultado do modelo e então entregue ao usuário final (ou a outra aplicação) esse resultado.

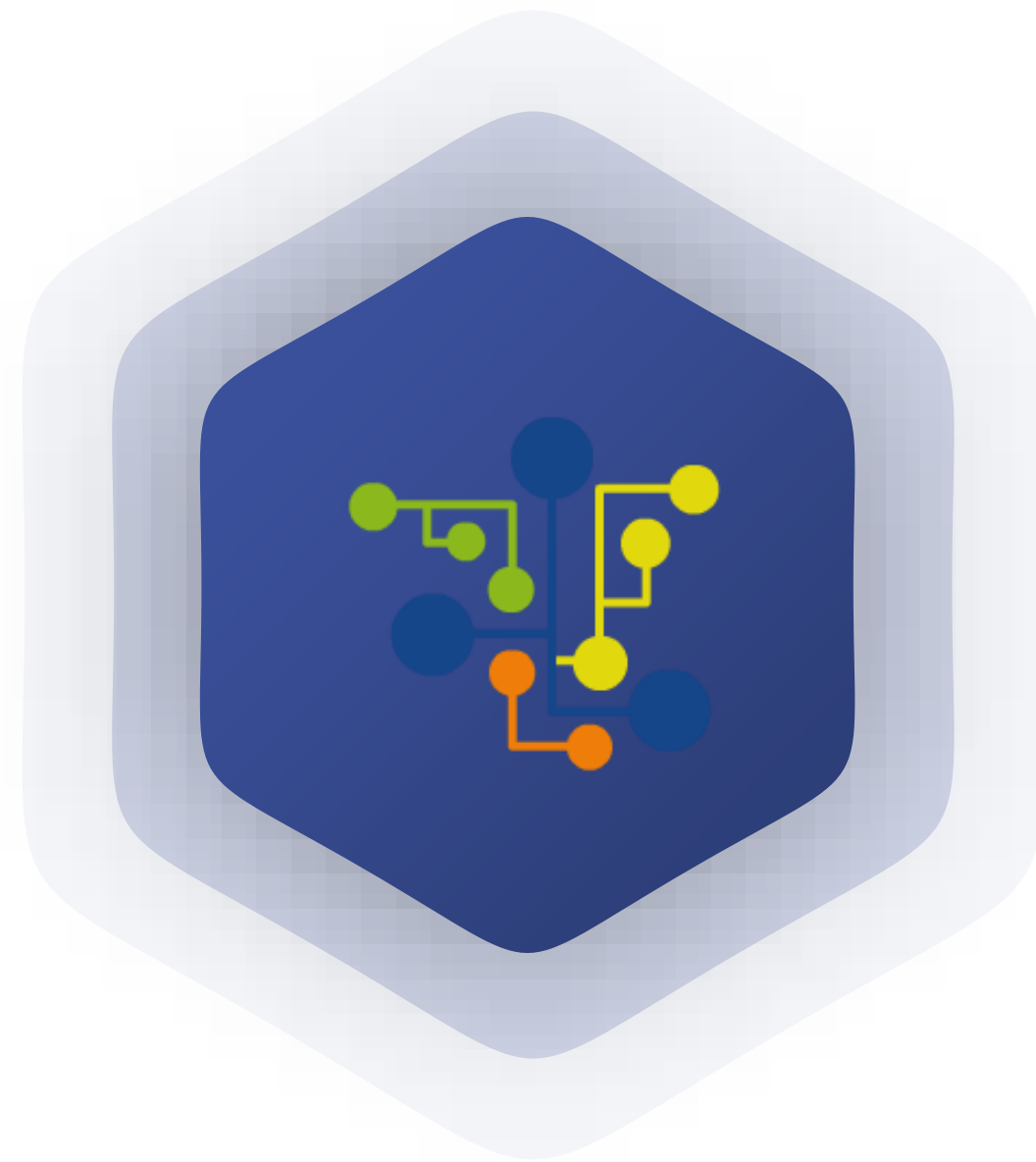


# Como Publicar Modelos de Machine Learning em Produção

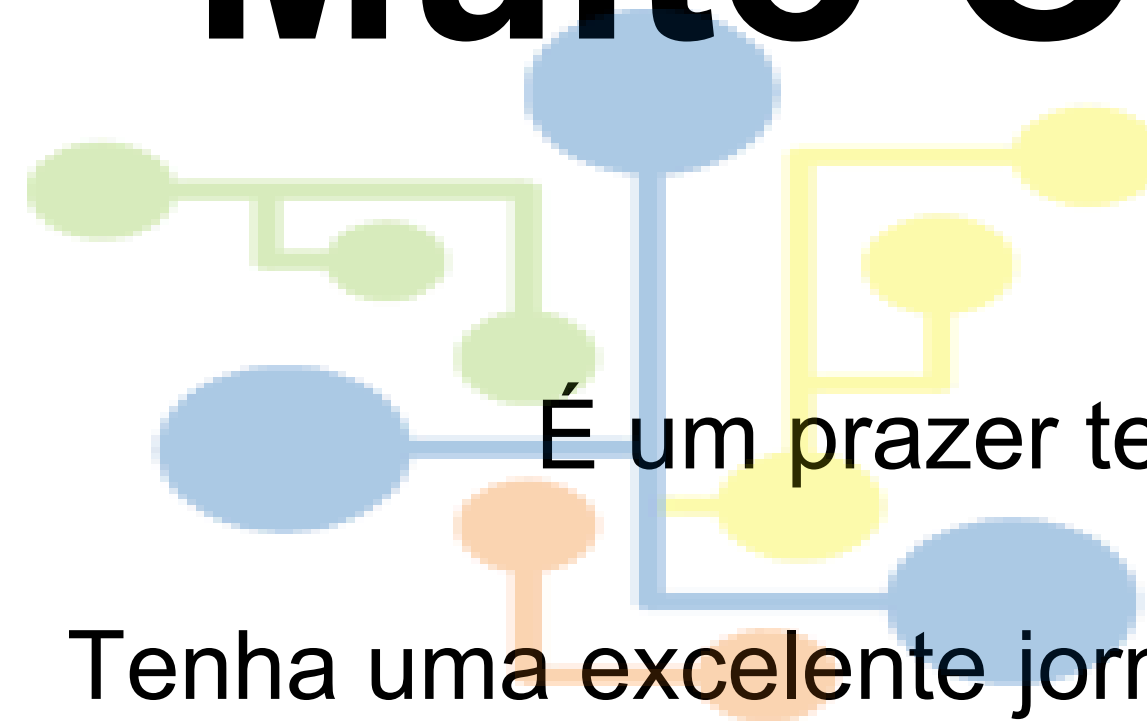
E como podemos fazer o Deploy de um modelo de Machine Learning?

- 1- Deploy do Modelo em nuvem com Google Cloud Platform, AWS ou Azure.
- 2- Aplicação Web com o Framework Web Flask (para linguagem Python), por exemplo.
- 3- Deploy do Modelo com TensorFlow Serving.
- 4- Criar uma REST API para nosso modelo já treinado.
- 5- Desenvolver uma aplicação em qualquer outra linguagem e usar JSON para entregar os resultados das previsões.





# Muito Obrigado.



É um prazer ter você aqui.

Tenha uma excelente jornada de aprendizagem.