
Deploy em Nuvem de Modelo de ML Utilizando a IBM Cloud

— Projeto Final de Computação em Nuvem —

Mariama Laianna Liviany Lucas Priscilla

Roteiro da Apresentação

1. Introdução;
2. Estado da Arte;
3. Estudo de Caso;
4. Considerações Finais.

Introdução

Introdução

- Computação em nuvem:
 - Utiliza hardware e software como recursos de serviços habilitados para virtualização.
- IBM, Amazon Web Services (AWS), Microsoft Azure:
 - Implantam os aplicativos através de uma rede virtual;
 - Oferece vários recursos com pouco capital inicial e custos operacionais.
- Objetivos:
 - Analisar os problemas relacionados ao deployment;
 - Avaliar o desempenho do modelo de machine learning em cloud.

Proposta

- Nosso trabalho propõe testar duas abordagens:
 - Inserir o modelo no container utilizando o Kubernetes da IBM Cloud e o docker;
 - Enviar o modelo utilizando a ferramenta de deploy da IBM cloud.

Estado da Arte

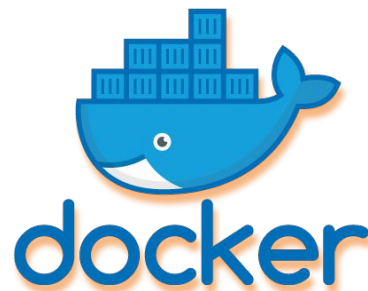
Computação em Nuvem

- Utilização sobre demanda;
- Conjunto de rede de alta escalabilidade;
- Qualidade de serviço;
- Infraestrutura barata;
- Acesso simples.



Docker

- Alternativa de virtualização mais leve;
- Implantação facilitada, rápida e aproveitamento dos recursos;
- Docker Registry.



Kubernetes

- Frameworks de orquestração de containers;
- Gerenciamento de aplicativos de forma automática e escalar;
- Arquitetura de comunicação entre os hosts mestre e escravos.



kubernetes

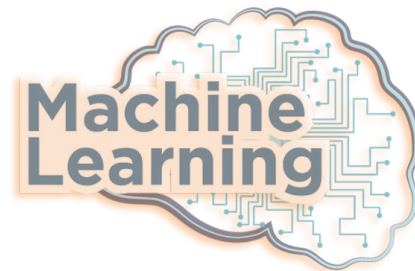
IBM Cloud

- Mais de 170 serviços;
- Desenvolver, implantar, executar e gerenciar aplicações.



Machine Learning

- Área da IA;
- Aprende através da experiência;
- Facilita a tomada de decisão.



Estudo de Caso

Deploy em Watson Machine Learning

- Primeiro utilizamos a plataforma Watson Machine Learning do IBM Cloud com o plano Lite (grátis) para realizar o deploy de um modelo de machine learning.
- Após realizar o deploy do modelo podemos fazer uma requisição por meio de API para que o modelo hospedado faça as predições e retorne os resultados.

Deploy em Container

- Fazer o deploy de um container que contém uma API Rest no cluster do Kubernetes gerenciado pela IBM Cloud.



**IBM Cloud
Kubernetes Service**

Criando um Cluster Kubernetes

IBM Cloud

Procurar recursos e produtos...

Catálogo Docs Suporte Gerenciar

Catálogo /

Kubernetes Cluster

Autor: IBM • Documentos • Docs de API

Criar

Sobre

Detalhes do plano

Saiba mais sobre as diferenças entre os planos em nossos docs.

Plano de precificação

Grátis

Serviço de orquestração

Selecione o tipo de plataforma de contêiner e a versão para seu cluster. Para obter mais informações sobre as versões, incluindo links para as notas sobre a liberação da comunidade da plataforma de contêiner, consulte os docs.

Kubernetes

1.22.10

Detalhes do recurso

Nome do cluster

mycluster-nuvem

Grupo de recursos

Default

Resumo

Estados Unidos

Kubernetes Cluster

1

Nó do trabalhador

Grátis - 2 vCPUs e 4 GB de RAM

Grátis

Custo total estimado

Grátis/mês


Encargos adicionais para redes e largura de banda podem ser aplicadas. O total mensal real variará com a precificação em camadas. A estimativa não inclui os custos de integração.

Criar

Incluir na estimativa

Painel do Kubernetes no Site da IBM Cloud

☰ IBM Cloud

Kubernetes

Clusters

Container Registry

Catálogo do Helm

Reservas

Procurar recursos e produtos...





Q

Catálogo

Docs

Suporte

Gerenciar ▾




Clusters do Kubernetes

Grupo de recursos: Filtrar ... ▾



Localização: Filtrar ... ▾

Q Procurar



Criar cluster

+

Nome	Estado	Localização	Contagem de trabalhadores	Criado	Versão	Infraestrutura
mycluster-nuvem	 Normal	Amsterdã 03	1	Expira em 30 dias	 1.22.10_1553	Clássico

Itens por página: 25 ▾

1 – 1 de 1 item

1 ▾

Página 1 de 1

◀ ▶

Painel do Cluster no Site da IBM Cloud

IBM Cloud

Procurar recursos e produtos...

Q

Catálogo

Docs

Suporte

Gerenciar

Clusters /

mycluster-nuvem

Normal

Expira em 25 dias

Incluir tags

Auxílio

Painel do Kubernetes

Ações...

Visão Geral

Nós do trabalhador

Conjuntos de trabalhadores

DevOps Novo

Status do nó

1 de 1

Normal

Detalhes

Status de complemento

0 de 0

Normal

Detalhes

Status do mestre

Normal

Documentos

Status do Ingress

Desconhecido

Documentos

Detalhes

ID do cluster

ca1nfgff0pp4d05hm7g

Versão

1.22.10_1553

Infraestrutura

Clássica

Zonas

Milão 01

Criado

16/06/2022 15:34

Grupo de Recursos

Default

Cumprimento de segurança de imagem

Ativar

Funcionamento do nó

Total de nós: 1

Crítico 0%

Aviso 0%

Normal 100%

Pendente 0%

Detalhes do nó do trabalhador

Configurando o Ambiente do IBM CLI

- Login
 - `ibmcloud cr login`
- Plug-ins
 - `ibmcloud plugin install -all`

Configurando o Namespace

- Criação:
 - `ibmcloud cr namespace-add proj-nuvem`
- Listagem dos Namespaces:
 - `ibmcloud cr namespace-list v`

Configurando o Namespace

```
PS C:\Program Files\IBM\Cloud\bin> ibmcloud cr namespace-add proj-nuvem
No resource group is targeted. Therefore, the default resource group for the account ('Default') is targeted.

Adding namespace 'proj-nuvem' in resource group 'Default' for account Mariama Oliveira's Account in registry de.icr.io..
..

Successfully added namespace 'proj-nuvem'

OK
PS C:\Program Files\IBM\Cloud\bin> ibmcloud cr namespace-list -v
Listing namespaces for account 'Mariama Oliveira's Account' in registry 'de.icr.io'...

Namespace      Resource Group  Created
proj-nuvem     Default        47 seconds ago

OK
PS C:\Program Files\IBM\Cloud\bin>
```

Namespace no Site da IBM Cloud

☰ IBM Cloud

Container Registry

Iniciação rápida

Namespaces 1

Repositórios 0

Imagens 0

Lixeira 0

Configurações

Procurar recursos e produtos...

Catálogo Docs Suporte Gerenciar

📅 📊 🔔 👤

Namespaces


Local

Frankfurt

Grupo de Recursos: Filtrar ...

🔍 Procurar

🔄 ⚙️ Criar +

<input type="checkbox"/>	Nome	Grupo de Recursos	Contagem de repositório	Contagem de imagens	Política de retenção
▼ <input type="checkbox"/>	 proj-nuvem	Default	1	1	Reten todas as imagens

Itens por página: 25

1 – 1 de 1 item

1

Página 1 de 1

◀ ▶

Selecionando o Cluster no IBM CLI

- Configuração:
 - `ibmcloud ks cluster config -c calnfgff0pp4d05hm7g`

```
PS C:\Program Files\IBM\Cloud\bin> ibmcloud ks cluster config -c calnfgff0pp4d05hm7g
OK
The configuration for calnfgff0pp4d05hm7g was downloaded successfully.

Added context for calnfgff0pp4d05hm7g to the current kubeconfig file.
You can now execute 'kubectl' commands against your cluster. For example, run 'kubectl get nodes'.
If you are accessing the cluster for the first time, 'kubectl' commands might fail for a few seconds while RBAC synchronizes.

PS C:\Program Files\IBM\Cloud\bin> kubectl get nodes
NAME                STATUS    ROLES    AGE   VERSION
10.144.213.220      Ready    <none>    5m30s v1.22.10+IKS

PS C:\Program Files\IBM\Cloud\bin>
```

Cópia da Imagem

- Cópia:
 - `docker tag nuvem:latest de.icr.io/proj-nuvem/nuvem:latest`

```
C:\Users\mcso>docker tag nuvem:latest de.icr.io/proj-nuvem/nuvem:latest
```

Push da Imagem

- Upload:
 - `docker push de.icr.io/proj-nuvem/nuvem:latest`

```
PS C:\Program Files\IBM\Cloud\bin> docker push de.icr.io/proj-nuvem/nuvem:latest
The push refers to repository [de.icr.io/proj-nuvem/nuvem]
14bc3b5db2e2: Pushed
408346f32923: Pushed
2ef9611d247d: Pushed
b4a62f0deac1: Pushing [=====>] 189.6MB/201.7MB
3dde1e8b1316: Pushed
52f29e7bfaa2: Pushed
c6679bcc482d: Pushed
aedb4b362167: Pushed
285c8ba236e9: Pushed
13b045a1dfd2: Pushed
2fbabeba902e: Pushing [=====>] 144.3MB/528.7MB
ee509ed6e976: Pushing [=====>] 34.04MB/151.9MB
9177197c67d0: Pushing [====>] 1.222MB/18.95MB
7dbadf2b9bd8: Pushing [>] 107kB/10.69MB
e7597c345c2e: Waiting
```


Imagem no Site da IBM Cloud

- Listagem das Imagens:
 - `ibmcloud cr image-list`

```
PS C:\Program Files\IBM\Cloud\bin> ibmcloud cr image-list
Listing images...
```

Repository	Tag	Digest	Namespace	Created	Size	Security status
de.icr.io/proj-nuvem/nuvem	latest	8d65e0f10767	proj-nuvem	1 hour ago	436 MB	Scanning...

```
OK
```

Imagem no Site da IBM Cloud

IBM Cloud

Container Registry

Iniciação rápida

Namespaces 1

Repositórios 1

Imagens 1

Lixeira 0

Configurações

Procurar recursos e produtos...

Catálogo

Docs

Suporte

Gerenciar

📅

📊

🔔

👤

Imagens

Local

Frankfurt

Visualizar por: Compilação

Procurar

Criar +

<input type="checkbox"/>	Repository@digest	Tags	Tipo de manifest	Criado	Tamanho	Status de segurança
<input type="checkbox"/>	proj-nuvem/nuvem@sha256:8d65e0f10767...	latest	Docker	há 4 horas	436 MB	16 problemas

Itens por página: 25 1 – 1 de 1 item

1 Página 1 de 1

Deploy

- Gera:
 - `kubectl create deployment nuvem-deployment --image=de.icr.io/proj-nuvem/nuvem`
- Publica:
 - `kubectl expose deployment/nuvem-deployment --type=NodePort --port=8080 --name=nuvem-service --target-port=8080`

```
PS C:\Program Files\IBM\Cloud\bin> kubectl create deployment nuvem-deployment --image=de.icr.io/proj-nuvem/nuvem
deployment.apps/nuvem-deployment created
PS C:\Program Files\IBM\Cloud\bin> kubectl expose deployment/nuvem-deployment --type=NodePort --port=8080 --name=nuvem-s
service --target-port=8080
service/nuvem-service exposed
```

Informações

- NodePort:
 - `kubectl describe service nuvem-service`


```
PS C:\Program Files\IBM\Cloud\bin> kubectl describe service nuvem-service
Name:                nuvem-service
Namespace:           default
Labels:              app=nuvem-deployment
Annotations:         <none>
Selector:            app=nuvem-deployment
Type:               NodePort
IP Families:        <none>
IP:                 172.21.198.52
IPs:                172.21.198.52
Port:               <unset> 8080/TCP
TargetPort:         8080/TCP
NodePort:           <unset> 31284/TCP
Endpoints:          172.30.108.77:8080
Session Affinity:   None
External Traffic Policy: Cluster
Events:            <none>
PS C:\Program Files\IBM\Cloud\bin>
```

Informações


- IP Público:
 - `ibmcloud ks worker ls --cluster mycluster-nuvem`




```
PS C:\Program Files\IBM\Cloud\bin> ibmcloud ks worker ls --cluster mycluster-nuvem
OK
ID                                     Public IP      Private IP      Flavor  State  Status  Zone
e   Version
kube-calnf0pp4d05hm7g-myclusternu-default-000000d2  169.51.203.17  10.144.213.220  free   normal Ready  mil
01   1.22.10_1554
PS C:\Program Files\IBM\Cloud\bin>
```


Workload Status do Cluster


 **kubernetes**

default

 Search

 **Workloads**

Workloads 

Cron Jobs

Daemon Sets

Deployments


Jobs

Pods

Replica Sets

Replication Controllers


Stateful Sets


Service 

Ingresses

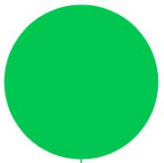
Services

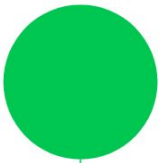
Config and Storage


Config Maps 

Persistent Volume Claims 


Workload Status


Running: 1
Deployments



Running: 1
Pods


Running: 1
Replica Sets


Deployments




Name	Namespace	Images	Labels	Pods
 nuvem-deployment	default	Show all	Show all	1 / 1


Deployments, Pods e Replica Sets


 **kubernetes**

default

 Search

 **Workloads**

Workloads 

Cron Jobs

Daemon Sets

Deployments


Jobs

Pods

Replica Sets

Replication Controllers


Stateful Sets


Service 

Ingresses

Services

Config and Storage

Config Maps 

Persistent Volume Claims 

Deployments

Name	Namespace	Images	Labels	Pods
 nuvem-deployment	default	Show all	Show all	1 / 1

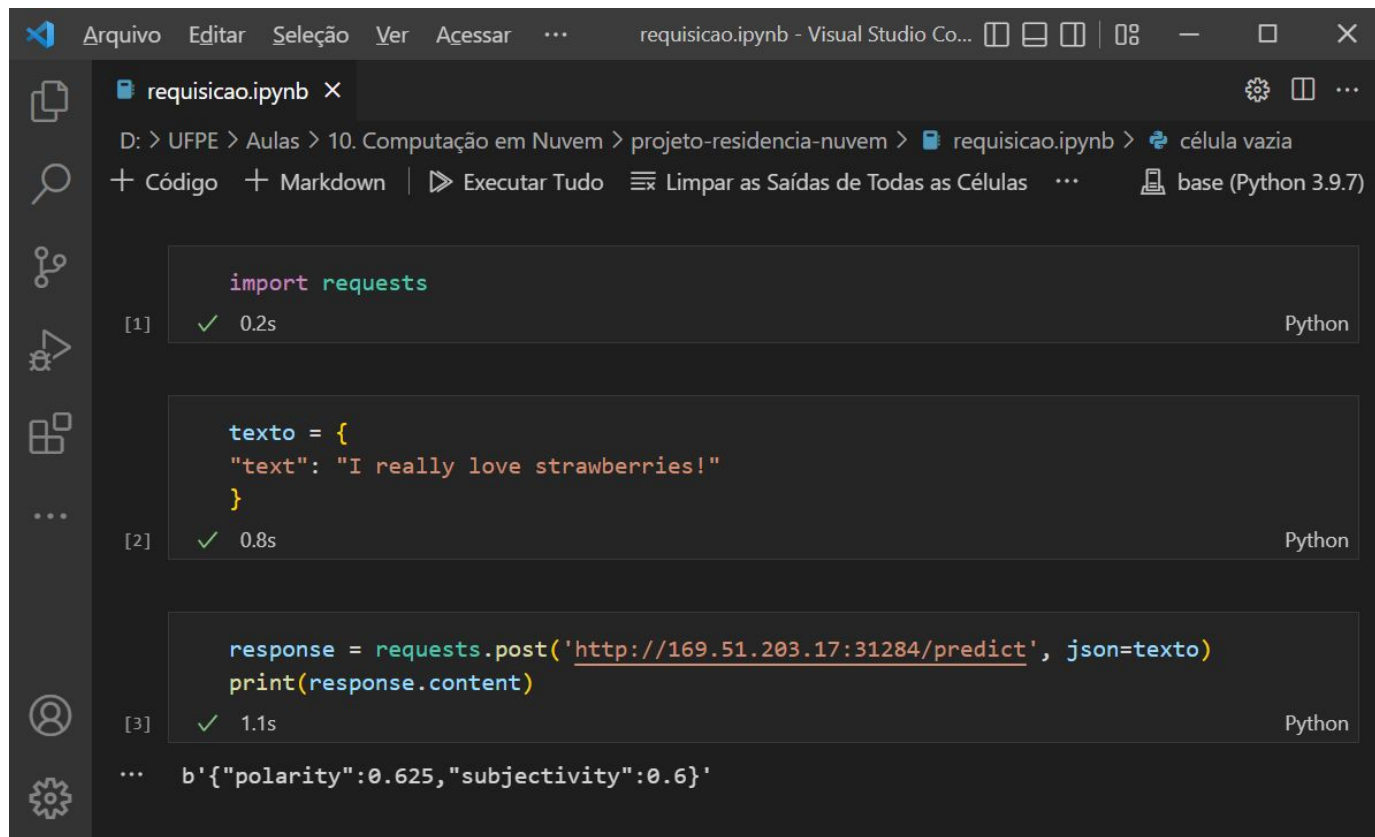
Pods

Name	Namespace	Images	Labels	Node	Status	Restarts	CPU Usage (cores)	Memory (bytes)
 nuvem-deployment-556984697f-j79wl	default	Show all	Show all	10.144.213.220	Running	0	<div>3.00m</div>	<div>19</div>

Replica Sets

Name	Namespace	Images	Labels	Pods
 nuvem-deployment-556984697f	default	Show all	Show all	1 / 1

Aplicação



requisicao.ipynb X

D: > UFPE > Aulas > 10. Computação em Nuvem > projeto-residencia-nuvem > requisicao.ipynb > célula vazia

+ Código + Markdown | ▶ Executar Tudo ≡ Limpar as Saídas de Todas as Células ... 🖨 base (Python 3.9.7)

```
[1] import requests
```

✓ 0.2s Python

```
[2] texto = {
    "text": "I really love strawberries!"
}
```

✓ 0.8s Python

```
[3] response = requests.post('http://169.51.203.17:31284/predict', json=texto)
    print(response.content)
```

✓ 1.1s Python

... b'{"polarity":0.625,"subjectivity":0.6}'

Aplicação

← → ↻ Não seguro | 169.51.203.17:31284/docs#/default/predict_predict_post 🔍 📄 ☆ 🔴 ⚙️ 🖨️ ☰ ⋮

FastAPI 0.1.0 OAS3

/openapi.json

default

POST /predict Predict

Parameters Cancel Reset

No parameters

Request body required application/json

```
{  
  "text": "I really love strawberries!"  
}
```

Execute Clear

Responses

Curl

```
curl -X 'POST' \  
  'http://169.51.203.17:31284/predict' \  
  -H 'accept: application/json' \  
  -H 'Content-Type: application/json' \  
  -d '{  
    "text": "I really love strawberries!"  
  }'
```

Request URL

```
http://169.51.203.17:31284/predict
```

Server response

Code	Details
200	<div>Response body<pre>{ "polarity": 0.625, "subjectivity": 0.6 }</pre>📄 Download</div>

Considerações Finais

Considerações Finais

- Tempo e facilidade de implementação;
- Manutenção;
- Monitoramento;
- Custo.

Tempo e Facilidade de Implementação

Deploy direto de modelo

- Tutoriais e documentação da IBM difíceis de serem seguidos;
- Documentação desatualizada;
- Tutoriais escritos e vídeos produzidos pela comunidade;
- Menos passos;
- Não aceita todos formatos.

Deploy de container

- Tutoriais e documentação da IBM difíceis de serem seguidos;
- Tutoriais escritos e vídeos produzidos pela comunidade → principalmente para configuração da API REST utilizando Docker;
- Vários passos que demandam diferentes ferramentas.

Manutenção

Deploy direto de modelo

- Maior abstração do sistema;
- Maior dificuldade em migrar o modelo para outra plataforma.

Deploy de container

- Menor abstração do sistema;
- Necessário manutenção do cluster e containers;
- Maior facilidade de realizar a migração.

Monitoramento

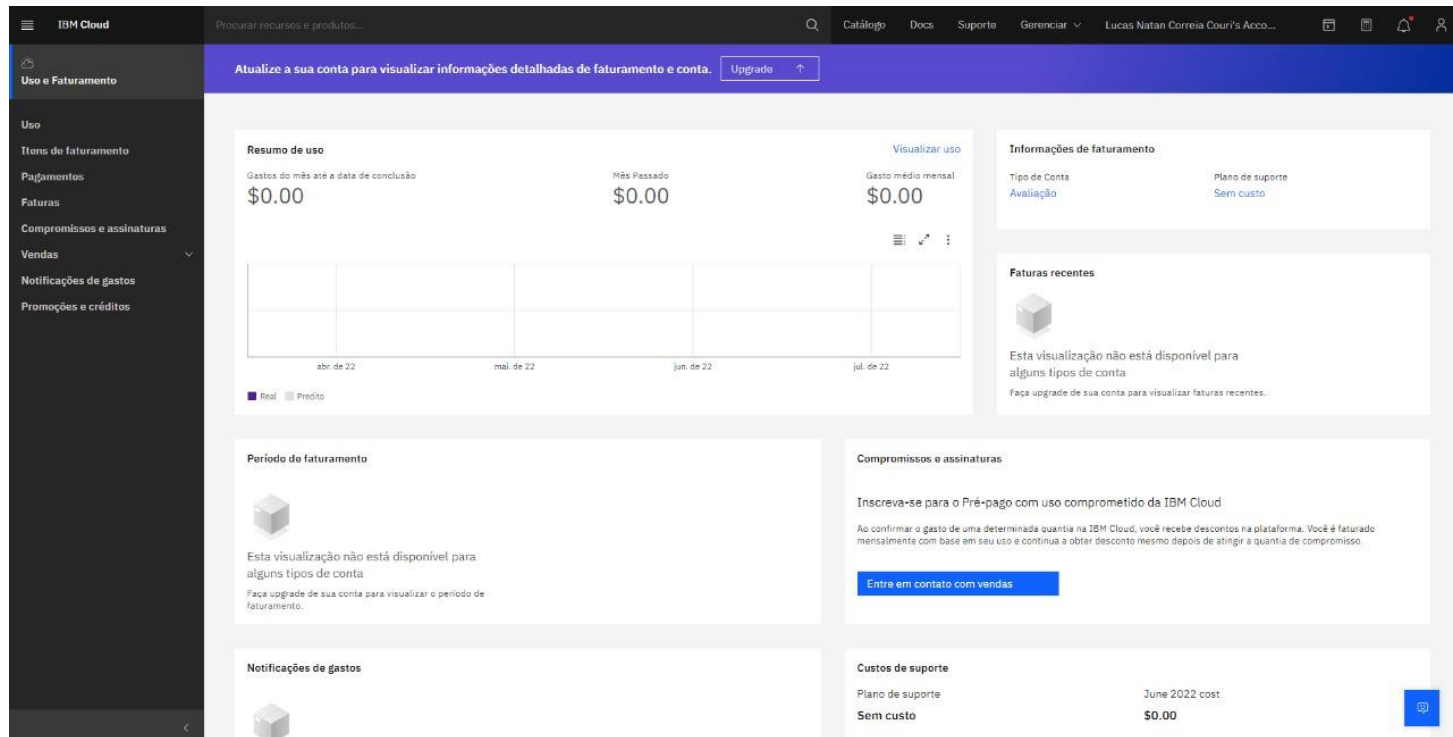
Deploy direto de modelo:

- Especificações sobre o modelo;
- Dashboard de monitoramento;
- Detecção de data drifts;
- **Isso tudo seria ótimo se a gente conseguisse achar isso na plataforma!**

Deploy de container:

- Infraestrutura de monitoramento a cargo do cientista de dados/desenvolvedor.

Custo



Custo

Deploy direto de modelo:

- Pagamento por hora de processamento na plataforma.

Deploy de container:

- Paga sempre para deixar a API REST disponível;
- Investimento inicial bastante alto, mas que pode valer a pena caso o projeto utilize todas as capacidades do Kubernetes.

Considerações Finais

- Os dois métodos foram difíceis de implementar devido a documentação ruim;
- Ter uma comunidade ativa utilizando a plataforma faz diferença;
- A escolha do método irá depender das necessidades do projeto.

Referências

Externas:

- [Kubernetes](#)
- [IBM Cloud For Education](#)
- [IBM Cloud Developer Tools CLI](#)
- [Getting started with IBM Cloud Container Registry](#)
- [Installing the tools and plug-ins manually](#)
- [Getting started with the IBM Cloud CLI](#)

Ícones:

- [Nuvem](#)
- [Docker](#)
- [Kubernetes](#)
- [Machine Learning](#)
- [IBM Cloud](#)

Obrigado!

Mariama Laianna Liviany Lucas Priscilla
