

# Data Master Case de Eng. de ML

MARCIO DE LIMA - JAN/22



# Contexto de Negócio

O diagnóstico de doenças cardíacas é feito com base em uma combinação de sinais clínicos e resultados de testes. Os tipos de testes solicitados variam de eletrocardiogramas, tomografia computadorizada (TC) cardíaca, exames de sangue e testes de esforço entre outros.

A definição de doença cardíaca é "...o que acontece quando o suprimento de sangue do seu coração é bloqueado ou interrompido por um acúmulo de substâncias gordurosas nas artérias coronárias...".



# Dados Estatísticos

As doenças cardíacas são a principal causa de morte no mundo: mais pessoas morrem anualmente por essas enfermidades do que por qualquer outra causa.

Estima-se que 17,9 milhões de pessoas morreram por doenças cardiovasculares em 2016, representando 31% de todas as mortes em nível global. Destes óbitos, estima-se que 85% ocorrem devido a ataques cardíacos e acidentes vasculares cerebrais (AVCs).

Mais de três quartos das mortes por doenças cardiovasculares ocorrem em países de baixa e média renda.

Das 17 milhões de mortes prematuras (pessoas com menos de 70 anos) por doenças crônicas não transmissíveis, 82% acontecem em países de baixa e média renda e 37% são causadas por doenças cardiovasculares.

– Fonte: <https://www.paho.org/pt/topicos/doencas-cardiovasculares>



# Fatores de Risco

Os fatores de risco para doenças cardíacas são:

- Colesterol alto;
- Pressão alta;
- Diabetes;
- Peso;
- Histórico familiar;
- Tabagismo

– Fonte: <https://www.bhf.org.uk/information-support/risk-factors>

# Fonte e Dados utilizados

- Fonte:

EWhat Causes Heart Disease? Explaining the Model

Link: <https://www.kaggle.com/tentotheminus9/what-causes-heart-disease-explaining-the-model#Diagnosing-Heart-Disease> Acesso em 26/12/2021

Autor: ROB HARRAND

Data: 3 anos atrás - 2018

Versão: 14

- Dados:

Heart Disease UCI

<https://archive.ics.uci.edu/ml/datasets/Heart+Disease>





# Dicionário de Dados

- Database: Cleveland – 303 instâncias; Missing Values: Sim; Dados Balanceados: Sim;
- age: The person's age in years
- sex: The person's sex (1 = male, 0 = female)
- cp: The chest pain experienced (Value 1: typical angina, Value 2: atypical angina, Value 3: non-anginal pain, Value 4: asymptomatic)
- trestbps: The person's resting blood pressure (mm Hg on admission to the hospital)
- chol: The person's cholesterol measurement in mg/dl
- fbs: The person's fasting blood sugar (> 120 mg/dl, 1 = true; 0 = false)
- restecg: Resting electrocardiographic measurement (0 = normal, 1 = having ST-T wave abnormality, 2 = showing probable or definite left ventricular hypertrophy by Estes' criteria)
- thalach: The person's maximum heart rate achieved
- exang: Exercise induced angina (1 = yes; 0 = no)
- oldpeak: ST depression induced by exercise relative to rest ('ST' relates to positions on the ECG plot. See more here)
- slope: the slope of the peak exercise ST segment (Value 1: upsloping, Value 2: flat, Value 3: downsloping)
- ca: The number of major vessels (0-3)
- thal: A blood disorder called thalassemia (3 = normal; 6 = fixed defect; 7 = reversable defect)
- **target: Heart disease (0 = no, 1 = yes)**

# Modelo

## Tipo

Modelo de Aprendizado Supervisionado de Classificação Binária.

## Algoritmo

RandomForestClassifier

**Resultado** (acurácia)

Treinamento: 92,15%

Teste: 81,63%

## Parâmetros utilizados – Versão original

```
RandomForestClassifier(bootstrap=True, class_weight=None, criterion='gini',  
                        max_depth=5, max_features='auto', max_leaf_nodes=None,  
                        min_impurity_decrease=0.0, min_impurity_split=None,  
                        min_samples_leaf=1, min_samples_split=2,  
                        min_weight_fraction_leaf=0.0, n_estimators=10, n_jobs=None,  
                        oob_score=False, random_state=None, verbose=0,  
                        warm_start=False)
```



# Modelo - Tunning

## Contexto e Escopo

No notebook entregue pelo Data Science, linguagem python, não consta os experimentos de outros modelos nem nenhum hyper tuning no modelo escolhido. Desta forma, o Tunning efetuado faz um experimento de outros modelos e no modelo escolhido é feito o hyper tuning cujo objetivo é ter uma performance melhor no modelo. Essa performance não necessariamente é ter somente um aumento de acurária e sim melhor generalização dos acertos das 2 classes do modelo.

## Ferramentas e linguagens

Utilizamos linguagem python 3.8, para os experimentos e validações dos parâmetros a api GridSearchCV e fizemos experimentos de normalização nos dados com a api MinMaxScaler, devido as diferenças de escalas entre as colunas.

Para a criação dos fontes foi utilizado o Jupyter Lab e foi utilizado um ambiente local (notebook Core i5 com 16GB Ram).



# Modelo - Tunning

## Testes com outros modelos e com normalização

\* Normalização

Algoritmo	Versão	Acurácia	Acurária ROC	Acurácia *	Acurária ROC *
XGBClassifier (tst)	4	81.97%	-	-	-
Logistic Regression	3	83.95%	92.33%	82.68%	90.56%
RandomForestClassifier	3	84.27%	89.03%	81.80%	91.42%
XGBClassifier	3	80.60%	88.34%	82.63%	88.64%
Gaussian NB	3	78.50%	86.95%	81.33%	88.32%
Decision Tree Classifier	3	76.03%	76.38%	74.35%	70.44%
KNN	3	62.35%	67.81%	80.55%	86.55%
SVM	3	57.45%	57.19%	78.47%	86.69%
RandomForestClassifier (tst)	2	-	-	81.97%	-
RandomForestClassifier (tst)	1	81.97%	-	-	-



# Modelo - Tunning

## Conclusão do tuning

O Tuning de versão 1 demonstrou melhor acurácia e maior acertos nas 2 targets (0 e 1). Tivemos um aumento de 5% na acurácia do treinamento e o mesmo resultado no teste, mas pela métrica de matriz de confusão e relatório de classificação o acerto entre as classes foi equalizado, tornando o algoritmo mais genérico. Não houve diferença significativa com a aplicação de normalização no dataset, desta forma, foi ignorada. O modelo XGBClassifier aparece como promissor, mas para o case, vamos seguir com a decisão do Data Science (Autor) com o RandomForestClassifier, já que a diferença foi muito pequena.

**Resultado** (acurácia)

Treinamento: 97.93%

Teste: 81.97%

## Melhores parâmetros

```
RandomForestClassifier(bootstrap=True, class_weight='balanced', criterion='gini', max_depth=10, max_features='auto',  
                        max_leaf_nodes=None, min_impurity_decrease=0.0, min_impurity_split=None, min_samples_leaf=3,  
                        min_samples_split=2, min_weight_fraction_leaf=0.0, n_estimators=125, n_jobs=None, oob_score=False,  
                        random_state=7, verbose=0, warm_start=False)
```

```
best params: {'max_depth': 10, 'min_samples_leaf': 3, 'min_samples_split': 2, 'n_estimators': 125}
```

```
best score: 0.9244156669776504
```



# Objetivo e Escopo – Case 1 (API)

## Objetivo

Através dos dados informados (13 informações) indicar se o paciente tem ou não a probabilidade de ter uma doença cardíaca.

## Uso

Ferramenta de simulação para médicos clínicos e/ou laboratórios para indicar um paciente para um tratamento especializado o mais rápido possível.

## Escopo

Para o alcance do objetivo acima, foi criado 2 aplicações de arquitetura de micro-serviços, conforme abaixo. As aplicações rodam em containers no AKS e seu build e deploy ocorrem via Azure DEVOPS.

- 1) Web-Site: Foi criado uma APP Web, onde o usuário possa digitar as informações necessárias e clicar no botão “Verificar” para obter o resultado; A página WEB exibirá um Sim junto com a probabilidade ou um Não;
- 2) App de back-End: Essa aplicação expõe o modelo de ML (pkl) do case como um Web-Service REST, tornando-o uma API para servir a APP Web.
- 3) A parte de infra-estrutura de Cloud (Azure) foi feita em Terraform (InfraAsCode)



# Objetivo e Escopo – Case 2 (Stream)

## Objetivo

Receber dados de pacientes via streaming e indicar se os pacientes têm ou não a probabilidade de ter uma doença cardíaca.

## Uso

Ferramenta on-line em streaming que recebe dados de exames de pacientes de laboratórios e armazena os resultados. Com os resultados, a ferramenta pode notificar o(s) médico(s) sobre o caso e proporcionar ao paciente um tratamento mais especializado e mais rápido possível.

## Escopo

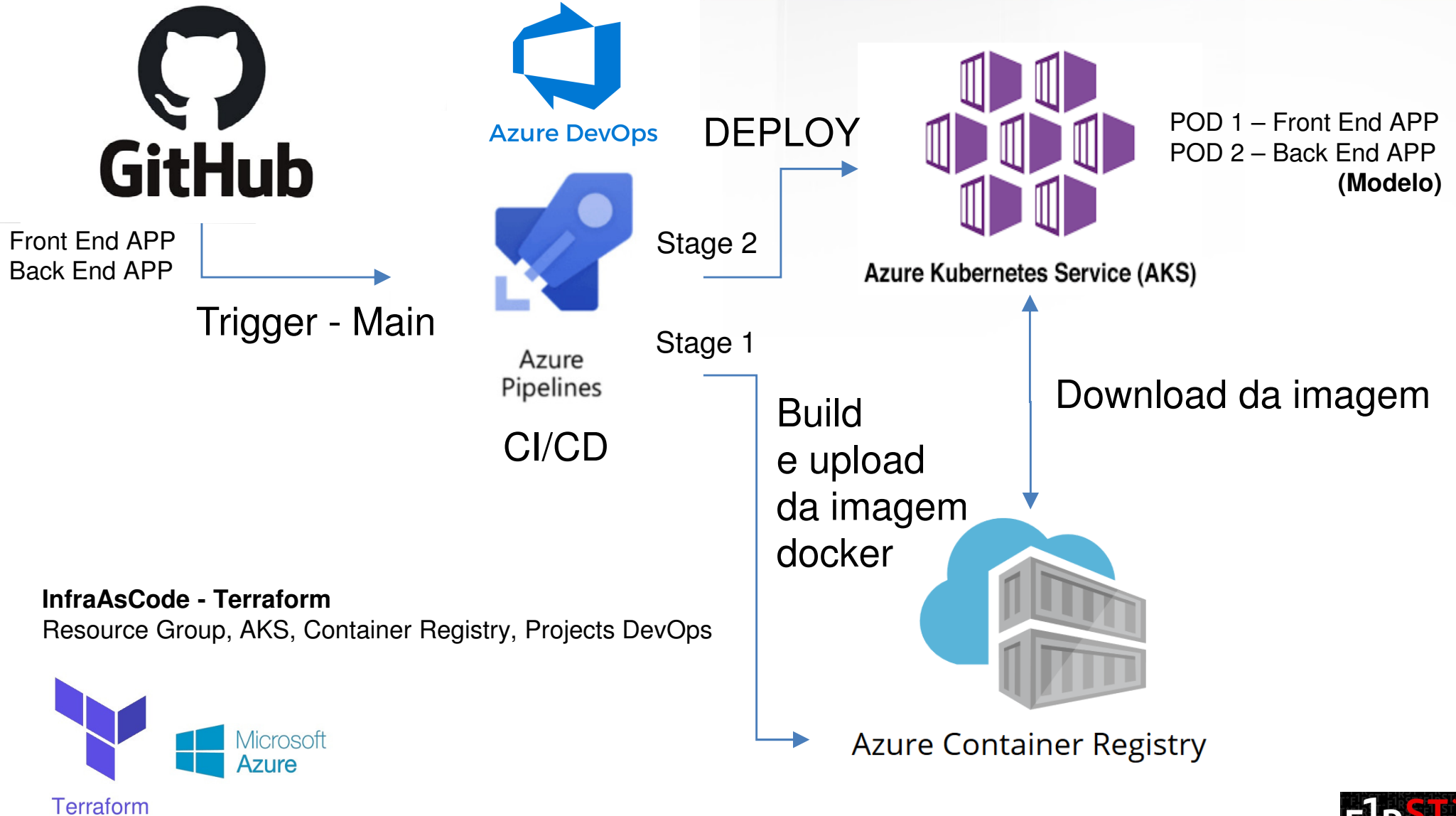
Para o alcance do objetivo acima, foi criado uma infra-estrutura em cloud (Azure) via script em Terraform para o processamento em streaming dos dados. Utilizado a plataforma Databricks, MLFLOW e o EventHub. Foi criado um notebook para o envio de dados simulando o envio por um laboratório e de um paciente. Após o envio dos dados, os mesmos são processados e registrados / gravados numa delta table do Databricks.

## Fora de Escopo

Notificação dos resultados ao(s) médico(s).

# Desenho de Solução

## Case 1 (API)

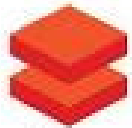




# Desenho de Solução Case 2 (Stream)

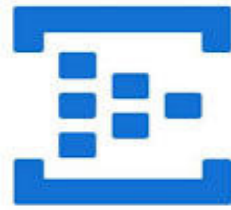


Fontes

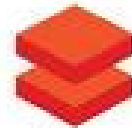


databricks

Gerador de Eventos  
Simulador dos laboratórios



Azure Event Hubs



databricks

Streaming  
Captura dos dados,  
preparação, aplicação  
do modelo e  
armazenamento do  
resultado

mlflow

Registry  
Registro dos modelos e  
versionamento

Experiments  
Registro com métricas

InfraAsCode - Terraform

Resource Group, Databricks Workspace, EventHub



Terraform



Microsoft  
Azure



Delta Table

# Evidências de Funcionamento Case 1 e 2 – Azure Cloud

The image displays two screenshots of the Microsoft Azure portal. The left screenshot shows the 'Recent' resources page, listing various subscriptions and resources. The right screenshot shows the 'Azure subscription 1' details page, which includes a warning about free credit expiration and several cost management charts.

**Recent Resources (Left Screenshot):**

- Refresh
- Feedback
- Assign tags
- Filter for any field...
- Showing 7 results.
- Name**
- Azure subscription 1
- databricks-case2
- case2StreamNamespace
- aks\_case1\_marcio
- aks\_case1\_marcio-agentpool
- acrCaseRestMarcio
- rsg\_case1\_md1

**Azure subscription 1 Details (Right Screenshot):**

**Warning:** Your remaining \$159.48 of free credit expires in 16 days. Upgrade to keep going with your account. →

**Spending rate and forecast:**

Line chart showing spending rate over time. Current cost: R\$230.06, Forecast: R\$630.86.

**Costs by resource:**

Donut chart showing costs by resource. Data:

Resource	Cost
aks-default-35305281-vmss	R\$82.64
acrCaseRestMarcio	R\$71.50
databricks-case2	R\$14.70
Others	R\$61.22

**Top products by number of resources:**

Bar chart showing top products by number of resources.

**Azure Defender coverage:**

Section for Azure Defender coverage.

**Top free services by usage:**

Service	Usage
Storage, Files, LRS Write Operations	0.89 / 1 (10K)
Storage, Standard HDD Managed Disks, Disk Operations	137.24 / 200 (10K)
Load Balancer, Standard, Data Processed	4.06 / 15 (1 GB)
Storage, Files, LRS Data Stored	1.2 / 5 (1 GB/Mo)
Load Balancer, Standard, Included LB Rules and Outbound	185.35 / 750 (1 GB/Mo)


# Evidências de Funcionamento

## Case 1 (REST) – App Web

Avaliação de Doença Cardíaca x +

← → ↻ Não seguro | 20.84.4.123:5000 ☆ M Pausada Atualizar

### Previsão de Doença Cardíaca



Idade \* Idade - Exemplo: 40

Sexo \* Feminino

Dor no Peito \* Angina Típica

Pressão Arterial em repouso (mm Hg) \* Pressão Arterial mmHg - Exemplo: 128

Colesterol em mg/dl \* Colesterol mg/dl - Exemplo: 200

Glicemia em jejum \* Menor que 120 mg/dl

Eletrocardiograma em repouso \* Normal

Máxima Frequência Cardíaca por minuto \* Máx Freq Cardíaca - Exemplo: 148

Angina induzida por exercício Não

Depressão de ST induzida por exercício em relação ao repouso \* Depressão de ST - Exemplo: 2.3

Inclinação do segmento ST de pico em exercício \* Inclinação para cima

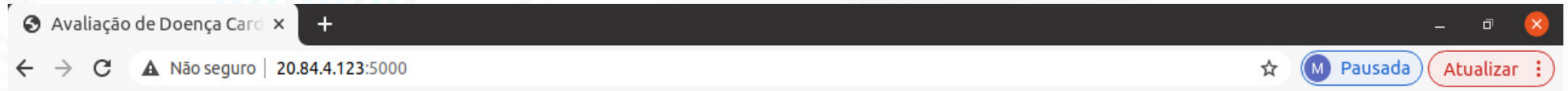
Artérias coronarianas em pleno funcionamento \* 0

Doença sanguínea chamada de talassemia \* Normal

Verificar

# Evidências de Funcionamento

## Case 1 (REST) - Preenchimento



## Previsão de Doença Cardíaca



Idade *	Sexo *	Dor no Peito *
<input type="text" value="70"/>	<input type="text" value="Masculino"/>	<input type="text" value="Angina Atípica"/>
Pressão Arterial em repouso (mm Hg) *	Colesterol em mg/dl *	Glicemia em jejum *
<input type="text" value="200"/>	<input type="text" value="400"/>	<input type="text" value="Maior que 120 mg/dl"/>
Eletrocardiograma em repouso *	Máxima Frequência Cardíaca por minuto *	Angina induzida por exercício
<input type="text" value="Hipertrofia ventricular esquerda provável ou definitiva"/>	<input type="text" value="300"/>	<input type="text" value="Sim"/>
Depressão de ST induzida por exercício em relação ao repouso *		
<input type="text" value="5.1"/>		
Inclinação do segmento ST de pico em exercício *	Artérias coronarianas em pleno funcionamento *	Doença sanguínea chamada de talassemia *
<input type="text" value="Inclinação para cima"/>	<input type="text" value="0"/>	<input type="text" value="Defeito Fixo"/>

Verificar



# Evidências de Funcionamento

## Case 1 (REST) - Resultado

### Previsão de Doença Cardíaca



Resultado: Encontrado indícios de doença

Probabilidade: 58.36%

Obs.: Atenção, esse resultado é uma previsão, sempre procure um médico cardiologista.

[Prever novo paciente](#)

### Previsão de Doença Cardíaca



Resultado: Não há indícios de doença

Obs.: Atenção, esse resultado é uma previsão, sempre procure um médico cardiologista.

[Prever novo paciente](#)



# Evidências de Funcionamento

## Case 1 (REST) - BackEnd

The screenshot displays the Postman application interface. At the top, there's a search bar and navigation icons. A yellow banner indicates the user is working locally in Scratch Pad. The main workspace shows a POST request named 'Chamar' to the endpoint 'http://52.151.229.54:8000/preverDoencaCard'. The request body is a JSON object with various patient data fields. The response is also in JSON format, showing a 'resultado' of 0.0 and a 'probabilidade' of 0.6073624696.

**Postman**

Search Postman

Working locally in Scratch Pad. [Switch to a Workspace](#)

POST Chamar

Modelo-BackEnd / Chamar

POST http://52.151.229.54:8000/preverDoencaCard

Send

Params Authorization Headers (8) **Body** Pre-request Script Tests Settings Cookies

none form-data x-www-form-urlencoded **raw** binary GraphQL JSON Beautify

```
1 { "pacienteId": "eeee", "age": 30.0, "sex": 1.0, "chest_pain_type": 4.0, "resting_blood_pressure": 128.0,
  "cholesterol": 200.0, "fasting_blood_sugar": 1.0, "rest_ecg": 0.0, "max_heart_rate_achieved": 200.0,
  "exercise_induced_angina": 0.0, "st_depression": 2.3, "st_slope": 2.0, "num_major_vessels": 3.0,
  "thalassemia": 3.0 }
2
```

Body Cookies Headers (4) Test Results

201 Created 659 ms 176 B Save Response

Pretty Raw Preview Visualize JSON

```
1 {
2   "resultado": 0.0,
3   "probabilidade": 0.6073624696
4 }
```

# Evidências de Funcionamento

## Case 1 (REST) - AKS e DevOps

portal.azure.com/#@9da468c6-2260-4466-9e2b-d0dbb15ef88b/resor ← → aex.dev.azure.com/me?mkt=pt-BR

Microsoft Azure Search resources, services, and docs (G+)

Home > Recent > aks\_case1\_marcio

### aks\_case1\_marcio | Workloads

Kubernetes service

Search (Ctrl+/) Add Delete Refresh Sh

Overview Activity log Access control (IAM) Tags Diagnose and solve problems Security

Kubernetes resources


- Namespaces
- Workloads
- Services and ingresses
- Storage
- Configuration

#### Deployments

Pods Replica sets Sta

Filter by deployment name Enter the full deployment name

<input type="checkbox"/>	Name				
<input type="checkbox"/>	coredns	kube-system	✓ 2/2	2	2 3 days
<input type="checkbox"/>	coredns-autoscaler	kube-system	✓ 1/1	1	1 3 days
<input type="checkbox"/>	metrics-server	kube-system	✓ 1/1	1	1 3 days
<input type="checkbox"/>	tunnelfront	kube-system	✓ 1/1	1	1 3 days
<input checked="" type="checkbox"/>	backendmodel	default	✓ 1/1	1	1 3 days
<input checked="" type="checkbox"/>	frontendmodel	default	✓ 1/1	1	1 3 days



**marciodelima1301** Editar perfil  
marciodelima1301@hotmail.com

Conta da Microsoft

Brasil  
marciodelima1301@hotmail.com

#### Organizações do Azure DevOps

dev.azure.com/marciodelima1301 (Proprietário)

##### Projetos

- case\_heart\_frontend
- case\_heart\_backend

Novo projeto

# Evidências de Funcionamento

## Case 1 (REST) - DevOps

dev.azure.com/marciodelima1301/case\_heart\_backend/\_build?definitionId=5

Azure DevOps

/ case\_heart\_backend / Pipelines / marciodelima.case1\_santa...

Search

case\_heart\_backend

Overview

Boards

Repos

Pipelines

Pipelines

marciodelima.case1\_santander\_engml\_backend

Runs Branches Analytics

Description

Stages

#20220115.3 Delete Pipeline\_Case1\_Backend\_MDL.json

Individual CI for main 3d8ad776

✓-✓

#20220115.2 Delete deployment directory

Individual CI for main 28504b36

✓-✓

dev.azure.com/marciodelima1301/case\_heart\_frontend/\_build?definitionId=6

Azure DevOps

/ case\_heart\_frontend / Pipelines / marciodelima.case1\_santa...

Search

case\_heart\_frontend

Overview

Boards

Repos

Pipelines

Pipelines

Environments

marciodelima.case1\_santander\_engml\_frontend

Runs Branches Analytics

Description

Stages

#20220115.2 Ajustes

Individual CI for main 72b91404

✓-✓

sábado

2m 21s

#20220115.1 Set up CI with Azure Pipelines

Individual CI for main c8547890

✓-✓

sábado

3m 24s

Edit Run pipeline

dev.azure.com/marciodelima1301/case\_heart\_frontend/\_build/results?buildId=29&view=results

Azure DevOps

/ Pipelines / marciodelima.case1\_santa... / 20220115.2

Search

#20220115.2 Ajustes

marciodelima.case1\_santander\_engml\_frontend

This run is being retained as one of 3 recent runs by pipeline.

Summary Environments Associated pipelines

Triggered by marciodelima

Repository and version  
marciodelima/case1\_santander\_engml\_frontend  
main 72b91404

Time started and elapsed  
sáb. at 18:27  
2m 21s

Related  
0 work items  
1 published

Stages Jobs

Build stage

1 job completed  
1 artifact

Deploy stage

1 job completed 37s

# Evidências de Funcionamento

## Case 2 (Stream) – EventHub e Databricks

portal.azure.com/#@9da468c6-2260-4466-9e2b-d0dbb15ef88b/resource/subscriptions/cb129d26-97d3-43a4-9782-7b2ca3e81c...

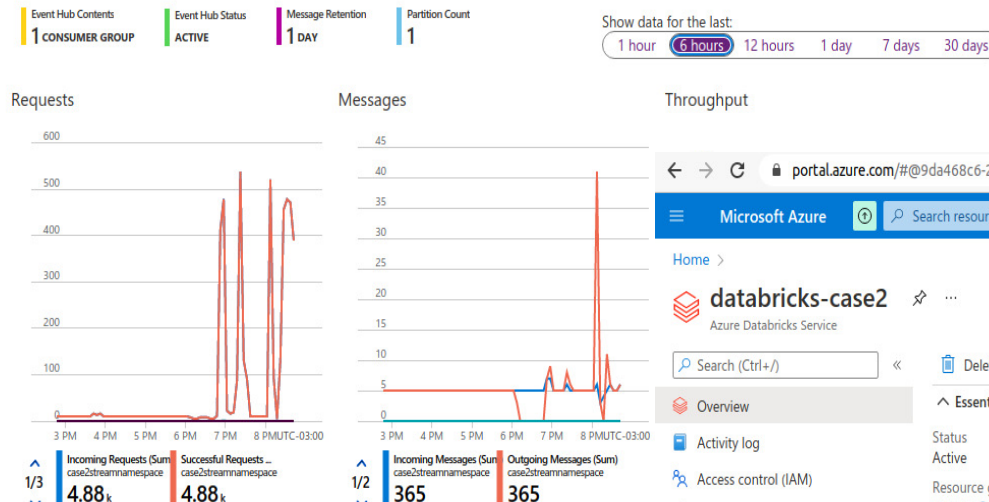
Microsoft Azure Upgrade Search resources, services, and docs (G+)

Home > case2StreamNamespace >

**dadoshearteventhub (case2StreamNamespace/dadoshearteventhub)** Event Hubs Instance

Search (Ctrl+/) Consumer group Delete Refresh

- Overview
- Access control (IAM)
- Diagnose and solve problems
- Settings
  - Shared access policies
  - Properties
  - Locks
- Entities
  - Consumer groups
- Features
  - Capture
  - Process data
- Automation
  - Task< (nrauiaw)



portal.azure.com/#@9da468c6-2260-4466-9e2b-d0dbb15ef88b/resource/subscriptions/cb129d26-97d3-43a4-9782-7b2ca3e81c...

Microsoft Azure Search resources, services, and docs (G+)

Home >

**databricks-case2** Azure Databricks Service

Search (Ctrl+/) Delete

Overview

- Activity log
- Access control (IAM)
- Tags

Settings

- Virtual Network Peering
- Encryption
- Properties
- Locks

Automation

- Tasks (preview)
- Export template

Support + troubleshooting

- New Support Request

Essentials

Status: Active

Managed Resource Group: databricks-rg-rsg\_case2\_md1

Resource group: rsg\_case2\_md1

URL: https://adb-343731794381320.0.azure.databricks.net

Location: Central US

Pricing Tier: standard

Subscription: Azure subscription 1

Subscription ID: cb129d26-97d3-43a4-9782-7b2ca3e81cd4

Tags (Edit)

Environment: Case 2 - Streaming





# Evidências de Funcionamento

## Case 2 (Stream) - Databricks

← → ↻ 🔒 adb-3437317794381320.azuredatabricks.net/?o=3437317794381320#setting/clusters

Microsoft Azure | Databricks

Portal marciodelima1301@hotmail.com

### Compute

All-purpose clusters Job clusters Pools

Create Cluster

All Created by me Accessible by me 🔍 Filter...

	Name	State	Nodes	Runtime	Driver	Worker	Creator		Actions
🌟	case2_cluster_stream_ml	Running	1	10.2 ML (includes Apache Spark 3.2.0, Sc...	Standard...	Standard...	marciodelim...	1	

1 1 of 1

← → ↻ 🔒 adb-3437317794381320.azuredatabricks.net/?o=3437317794381320#mlflow/experiments

Microsoft Azure | Databricks

### Repos

marciodelima1301@hotmail.com

case2\_santander... main

Fontes\_Databricks

Geral

mlruns

modelo

Add Repo

← → ↻ 🔒 adb-3437317794381320.azuredatabricks.net/?o=3437317794381320#job/list

Microsoft Azure | Databricks

Portal marciodelima1301@hotmail.com

### Jobs

Create Job

All Created by me 🔍 Filter

Name	Job ID	Created by	Task	Cluster	Schedule	Last run	Actions
Gerar_Eventos	77	marciodelima13...	gerarEventos	case2_cluster_stream_ml	Every minut...	Succeeded	
Capture_Eventos	84	marciodelima13...	capturaEventos	case2_cluster_stream_ml	Paused - Eve...	Succeeded	



# Evidências de Funcionamento

## Case 2 (Stream) – Databricks Jobs

← → ↻ 🔒 adb-3437317794381320.0.azuredatabricks.net/?o=3437317794381320#job/84/run/2/resultsOnly ☆ 🧑 Anônima Atualizar

Microsoft Azure | Databricks Portal marciodelima1301@hotmail.com

Jobs / Capture\_Eventos / Run 97005

### Capture\_Eventos run

Cancel job run

Output Show code Export as HTML Task run details

▼ f285cbab-2511-4e7f-a558-211a31f7088b Last updated: 15

Dashboard Raw Data

Input vs. Processing Rate  
records per second

0 rec/s 0 rec/s  
Input rate Processing rate

Batch Duration  
in milliseconds

2004.6 ms 61 ms  
Average Latest

Microsoft Azure | Databricks Portal marciodelima1301@hotmail.com

Jobs / Gerar\_Eventos

### Gerar\_Eventos

More Run now

Runs Task

Active runs Refresh

Start time	Run ID	Launched	Duration	Spark	Status
Jan 19 2022, 20:42 P...	97056	By scheduler	5s	Spark UI / Logs / Metrics	Running

Completed runs (past 60 days) Refresh

Latest successful run (refreshes automatically)

Start time	Run ID	Launched	Duration	Spark	Status
Jan 19 2022, 20:41 P...	97031	By scheduler	14s	Spark UI / Logs / Metrics	Succeeded
Jan 19 2022, 20:41 P...	96981	Manually	8s	Spark UI / Logs / Metrics	Succeeded
Jan 19 2022, 20:40 P...	96964	By scheduler	8s	Spark UI / Logs / Metrics	Succeeded

Job details

Job ID 77 🔗

Creator marciodelima1301@hotmail.com

Run as marciodelima1301@hotmail.com

Schedule

Every minute (UTC-03:00 — Brasília)

Edit schedule Pause

Cluster

case2\_cluster\_stream\_ml

Driver: Standard\_DS3\_v2, Workers: Standard\_DS3\_v2, 0 workers, 10.2 ML (includes Apache Spark 3.2.0, Scala 2.12)

View cluster Change Spark UI Logs

Metrics

# Evidências de Funcionamento

## Case 2 (Stream) – Dados

adb-3437317794381320.0.azuredatabricks.net/?o=34373177943

Microsoft Azure | Databricks

heart.dadospaciente

Refresh

Sample Data:

	angina	st_depression	st_slope	num_major_vessels	thalassemia	resultado	probabilidade_Nao	probabilidade_Sim	data
1		2.3	3	3	3	0	0.7071371365863812	0.292862863413619	2022-01-19
2		2.3	2	2	1	0	0.7439762968406025	0.25602370315939765	2022-01-19
3		2.3	1	2	3	0	0.7466984687845556	0.2533015312154444	2022-01-19
4		2.3	2	0	3	0	0.5441889914550755	0.4558110085449247	2022-01-19
5		2.3	3	2	2	0	0.6304289835067894	0.3695710164932105	2022-01-19
6		2.3	1	0	1	0	0.7356550326746969	0.264344967325303	2022-01-19
7									

Showing all 24 rows.

Schema:

	col_name	data_type	comment
1	patientid	string	
2	age	int	
3	sex	int	
4	chest_pain_type	int	
5	resting_blood_pressure	int	
6	cholesterol	int	
7	fasting_blood_sugar	int	

Showing all 21 rows.

# Evidências de Funcionamento

## Case 2 (Stream) - MLFlow

← → ↻ 🔒 adb-3437317794381320.0.azure.databricks.net/?o=3437317794381320#mlflow/models ☆ Anônima Atualizar

Microsoft Azure | Databricks Portal marciodelima1301@hotmail.com

### Registered Models

Share and serve machine learning models. [Learn more](#)

Create Model 🔍 Search by model name Search Filter Clear

Name	Latest Version	Staging	Production	Last Modified	Tags	Serving ?
Original	Version 1	—	Version 1	2022-01-15 20:52:28	—	—
Tunning	Version 5	Version 4	Version 5	2022-01-19 16:44:07	—	—

# Evidências de Funcionamento

## Case 2 (Stream) - MLFlow

Microsoft Azure | Databricks

### Tunning

[Details](#) [Serving](#)

Notify me about ⓘ All new activity ▼

Created Time: 2022-01-15 20:52:21      Last Modified: 2022-01-19 16:44:07      Creator: mai

► Description [Edit](#)

► Tags

▼ Versions    All    Active 4    [Compare](#)

<input type="checkbox"/>	Version	Registered at	Created by	Stage
<input type="checkbox"/>	<a href="#">Version 5</a>	2022-01-19 16:44:00	marciodelima1301@hotmail.com	<a href="#">Production</a>
<input type="checkbox"/>	<a href="#">Version 4</a>	2022-01-19 16:14:06	marciodelima1301@hotmail.com	<a href="#">Staging</a>
<input type="checkbox"/>	<a href="#">Version 3</a>	2022-01-19 10:49:28	marciodelima1301@hotmail.com	<a href="#">Staging</a>

## Evidências de Funcionamento Case 2 (Stream) – MLFlow Experimentos

Microsoft Azure | Databricks

Portal marc

### Experiments Preview

Create AutoML Experiment

Owned by me Accessible by me

Search experiments

Name	Location	Last Modified	Created by	Notes
<a href="#">Tunning_Version2_mlflow</a> /	/Repos/marciodelima1301@hotmail.com/case2_san...	2022-01-19 16:43:29 -03	marciodelima1301@ho...	
<a href="#">registrarModelos_pkl</a> /	/Repos/marciodelima1301@hotmail.com/case2_san...	2022-01-15 20:52:08 -03	marciodelima1301@ho...	



# Evidências de Funcionamento

## Case 2 (Stream) – MLFlow Experimentos

← → ↻ 🔒 adb-3437317794381320.0.azure.databricks.net/?o=3437317794381320#mlflow/experiments/1468857491982786 ☆ 📷 Anônima Atualizar

Microsoft Azure | Databricks Portal marciodelima1301@hotmail.co

/Repos/marciodelima1301@hotmail.com/case2\_santander\_engml\_stream/Fontes\_Databricks/Tunning\_Version2\_mlflow Preview

📘 Track machine learning training runs in an experiment. [Learn more](#) ✕

Experiment ID: 1468857491982786

► Description [Edit](#)

🔄 Refresh Compare Delete Download CSV ⬇️ ⌵ Start Time ⌵ All time ⌵

☰ ☒ Columns Only show differences ☐ 🔍 metrics.rmse < 1 and params.model = "tree" Search ⌵ Filter Clear

Showing 24 matching runs

								Metrics >		
<input type="checkbox"/>	↓ Start Time	Duration	Run Name	User	Source	Version	Models	auc	best score	best_cv_score
<input type="checkbox"/>	⊕ 9 hours ago	6.5h	-	marciodeli...	📄 Tunning_V	-	🔗 Tunning/5, 2 more	-	0.883	0.886
<input type="checkbox"/>	⊕ 20 hours ago		-	marciodeli...	📄 Tunning_V	-	🔗 Tunning/2, 2 more	-	0.886	0.886

Load more

The background of the slide features a stylized image of a document or report. It includes a line chart at the top left with two data series, one in blue and one in black, plotted against a grid. Below the line chart is a bar chart with blue bars. At the bottom left, there is a section titled 'Graph / Statistic' followed by some Latin placeholder text. The overall theme is data analysis and reporting.

# Fontes e Referências

## Case1 – Rest

- Geral  
([https://github.com/marciodelima/case1\\_santander\\_engml\\_geral.git](https://github.com/marciodelima/case1_santander_engml_geral.git) )
- Front-End  
([https://github.com/marciodelima/case1\\_santander\\_engml\\_frontend.git](https://github.com/marciodelima/case1_santander_engml_frontend.git) )
- Back-End  
([https://github.com/marciodelima/case1\\_santander\\_engml\\_backend.git](https://github.com/marciodelima/case1_santander_engml_backend.git) )

## Case2 – Stream

- Fontes  
([https://github.com/marciodelima/case2\\_santander\\_engml\\_stream.git](https://github.com/marciodelima/case2_santander_engml_stream.git) )



h / Statistic 2

Gráfico: Statistic 2

2017/18

2016/17

Dec

# Obrigado