

MBA em Inteligência Artificial

Big Data e Visualização de Dados

Prof. Diego Nogare



Faculdade de Computação e Informática
Mackenzie

Diego Nogare



Former MVP Artificial Intelligence
Microsoft Regional Director
Diretor no PASS.org
Chief Data Officer @ Lambda3
Membro notável na I2AI
Mestre em IA

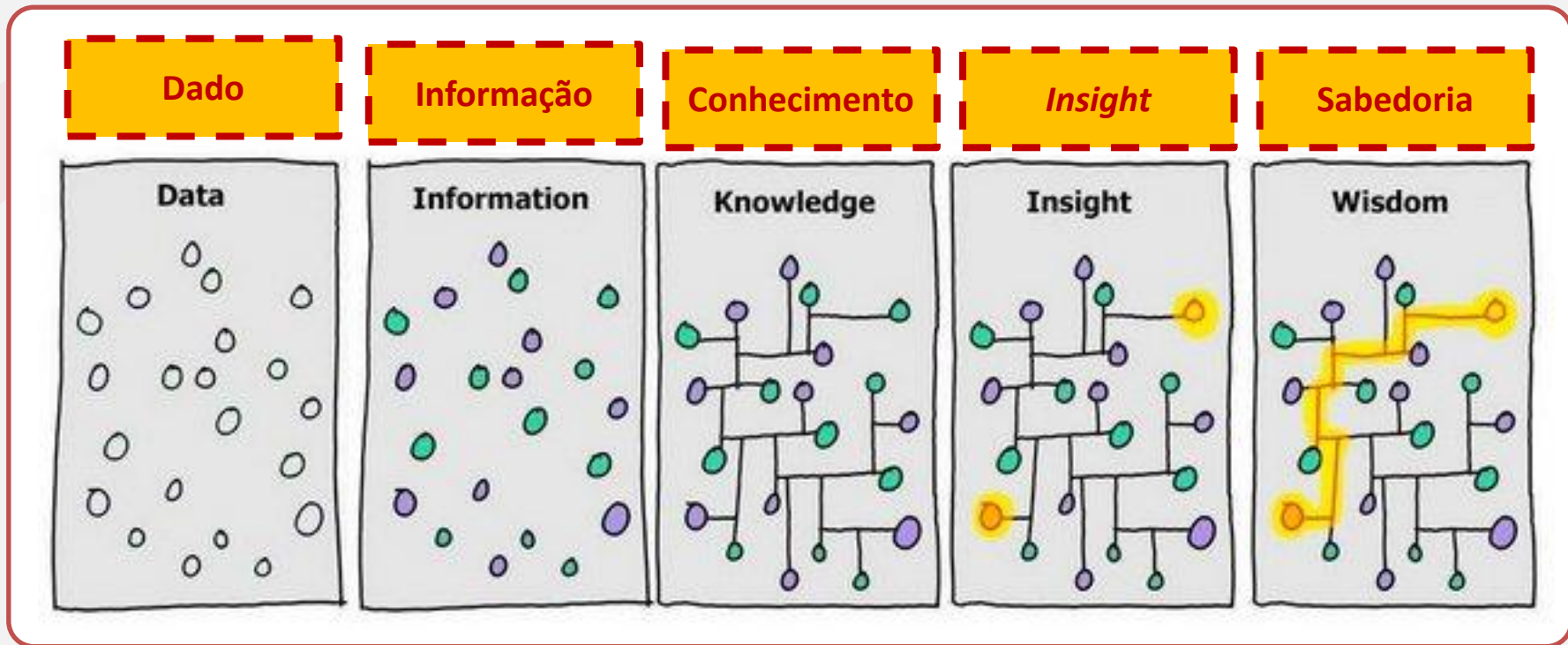
www.diegonogare.net
www.livrosdonogare.com.br





AQUISIÇÃO DO DADO

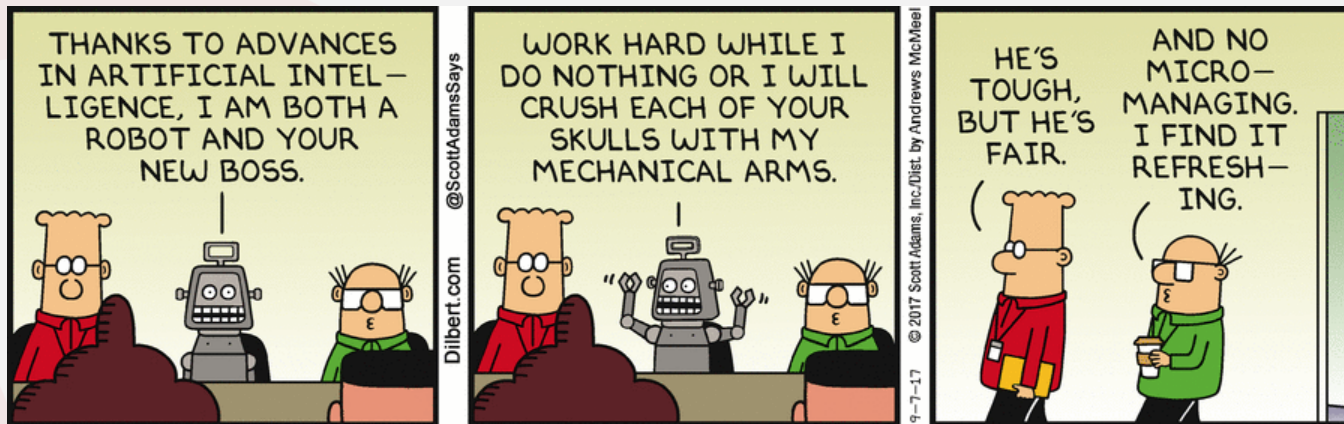
Descoberta do conhecimento



Machine Learning?!



Inteligência Artificial?!



Feliz pelos avanços em Inteligência Artificial, além de um robô também sou seu novo chefe.

Trabalhem pesado enquanto eu não faço nada ou vou esmagar seus crânios com meu braço mecânico.

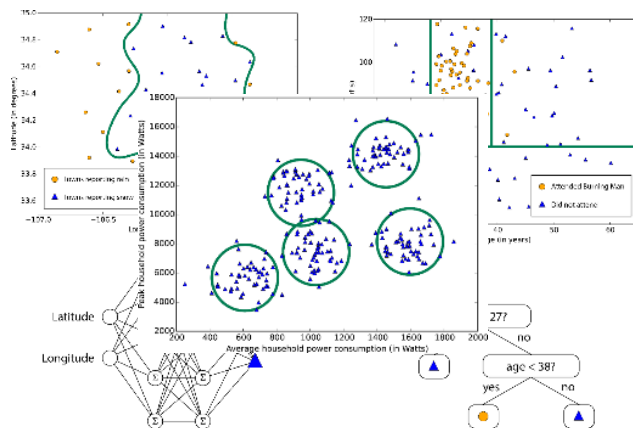
Ele é durão mas é justo.

E não faz micro-gereciamento. Isso é revigorante!

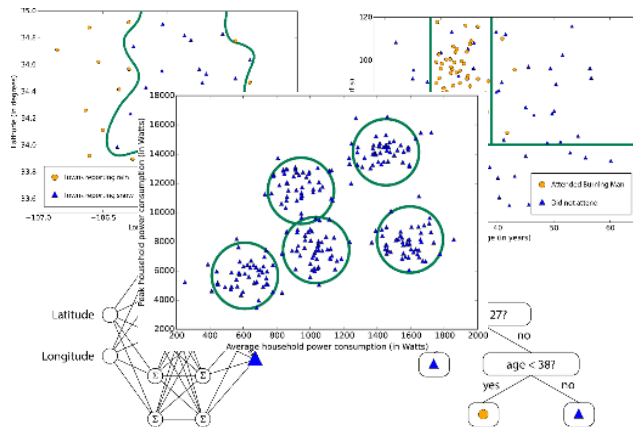
Aquisição de dados

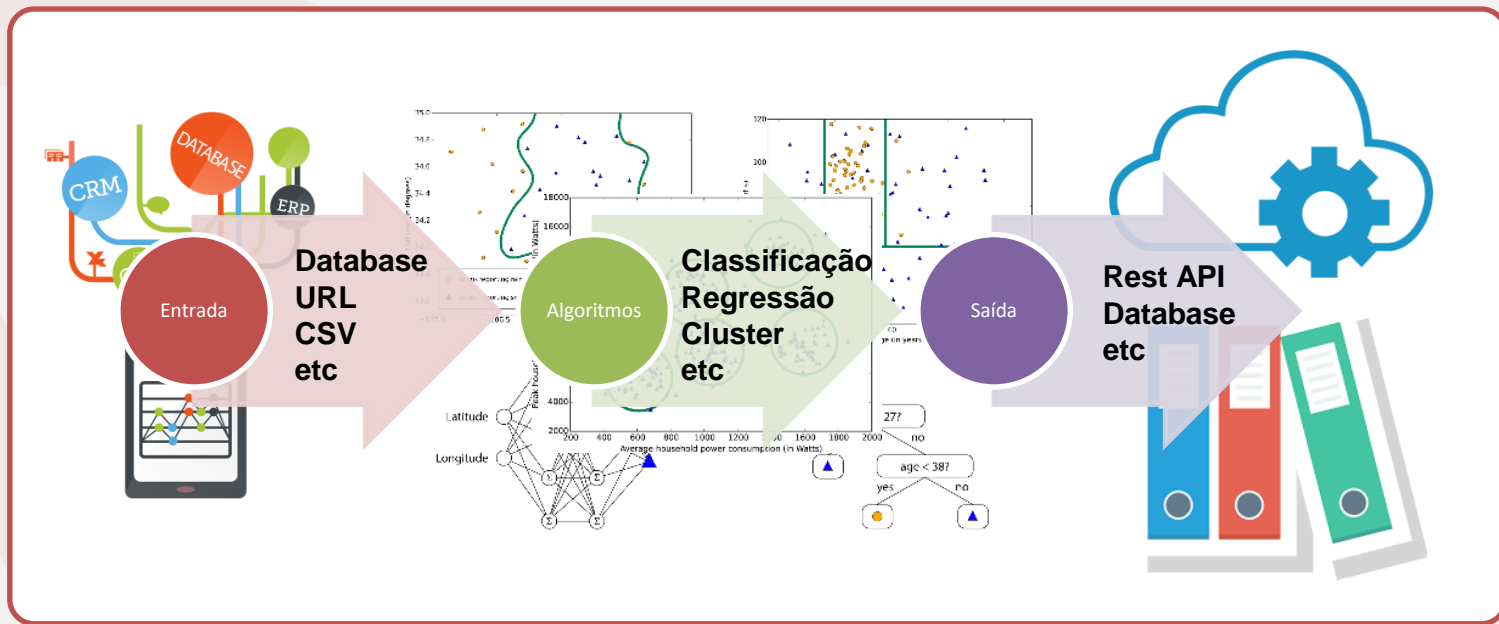


Criação do Modelo Preditivo

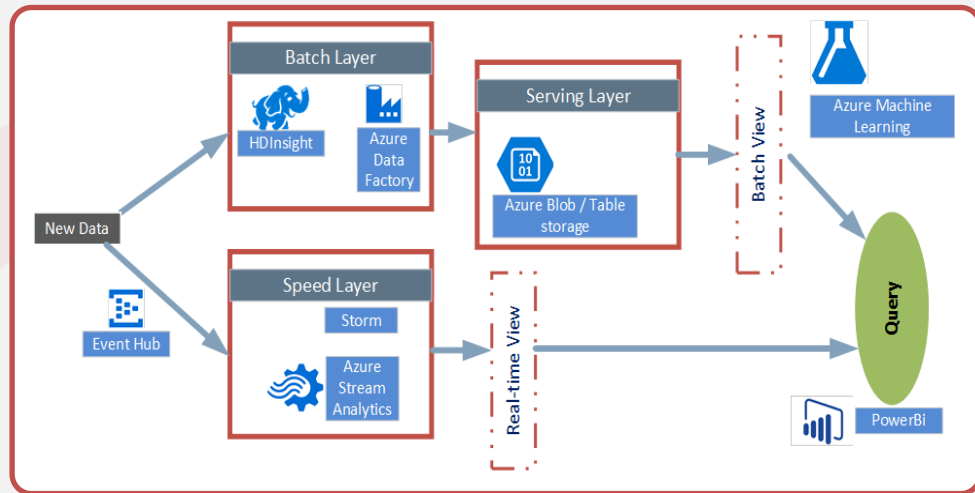


Saída do Processo





Implementação da Lambda Architecture



Speed Layer

A camada de velocidade processa fluxos de dados em tempo real. Esta camada sacrifica a taxa de transferência, pois visa minimizar a latência ao fornecer visões em tempo real.

Batch Layer

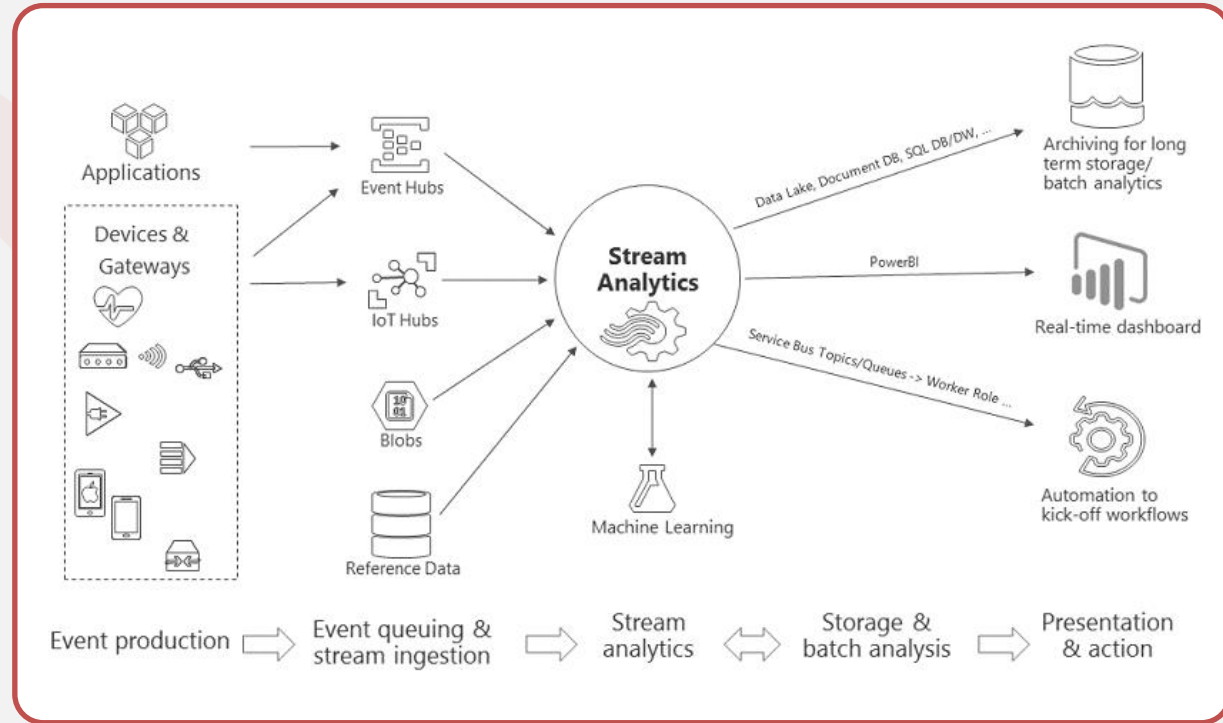
A Batch Layer compacta os resultados usando um sistema de processamento distribuído que pode lidar com grandes quantidades de dados

Serving Layer

A saída das camadas de Batch Layer e de Speed Layer é armazenada na camada de Serving Layer.

Stream Analytics

O Azure Stream Analytics é um mecanismo de processamento de eventos totalmente gerenciado que lhe permite configurar cálculos de análise em tempo real no fluxo de dados. Os dados podem vir de dispositivos, sensores, sites, feeds de mídia social, aplicativos...



Atividade com a base MTCars



	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
Fiat 128	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
Honda Civic	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
Toyota Corolla	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
Dodge Challenger	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
AMC Javelin	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
Camaro Z28	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
Pontiac Firebird	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
Fiat X1-9	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
Porsche 914-2	26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
Lotus Europa	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
Ford Pantera L	15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
Ferrari Dino	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
Maserati Bora	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8
Volvo 142E	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2

mpg - Miles/(US) gallon

cyl - Number of cylinders

disp - Displacement (cu.in.)

hp - Gross horsepower

drat - Rear axle ratio

wt - Weight (1000 lbs)

qsec - 1/4 mile time

vs - V/S

am - Transmission (0 automatic
1 manual)

gear - Number of forward gears

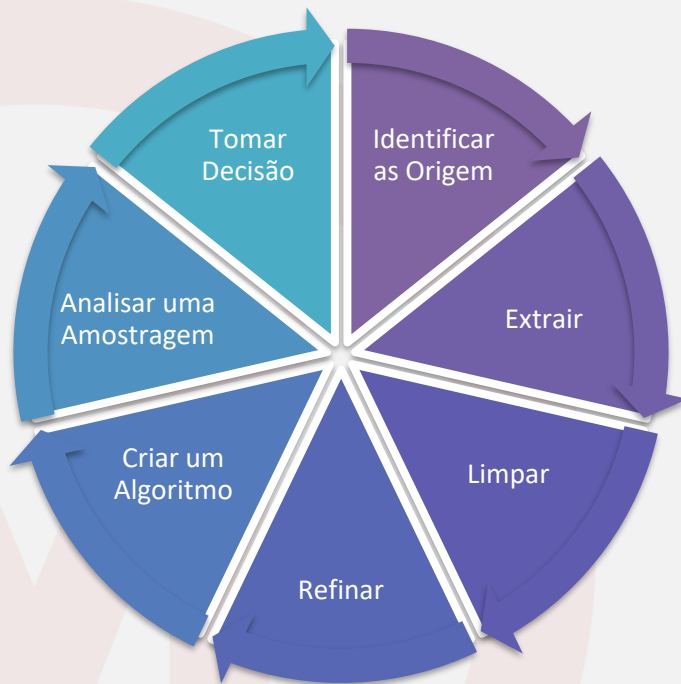
carb - Number of carburetors



ÁREAS DE ATUAÇÃO E ESPECIALIZAÇÃO DO PROFISSIONAL



■ Processo cíclico



1. Identificação do Problema

2. Coleta e limpeza de dados

**3. Desenvolvimento do
Modelo Preditivo**

4. Publicação do Modelo

**5. Monitoramento da
performance**







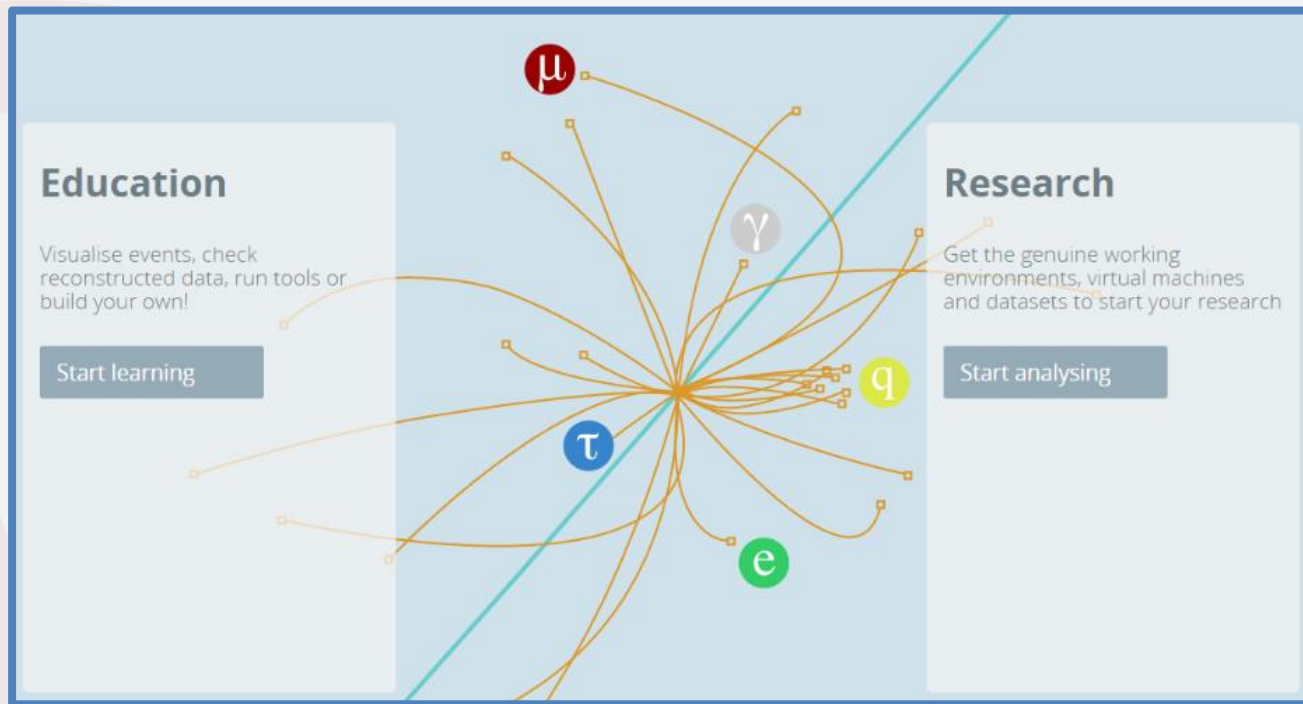
APLICAÇÕES NO MUNDO REAL

Áreas que já utilizam Big Data e Data Science em sua rotina

Dados abertos



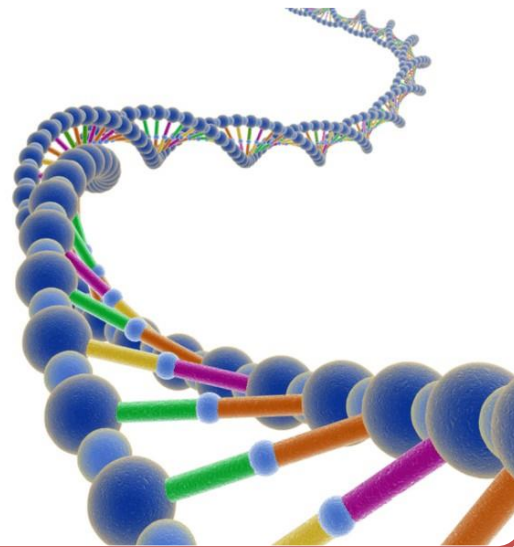
300 TB de dados abertos no CERN



Sequência do Genôma humano

1998 → 2001
2,8 Bilhões USD

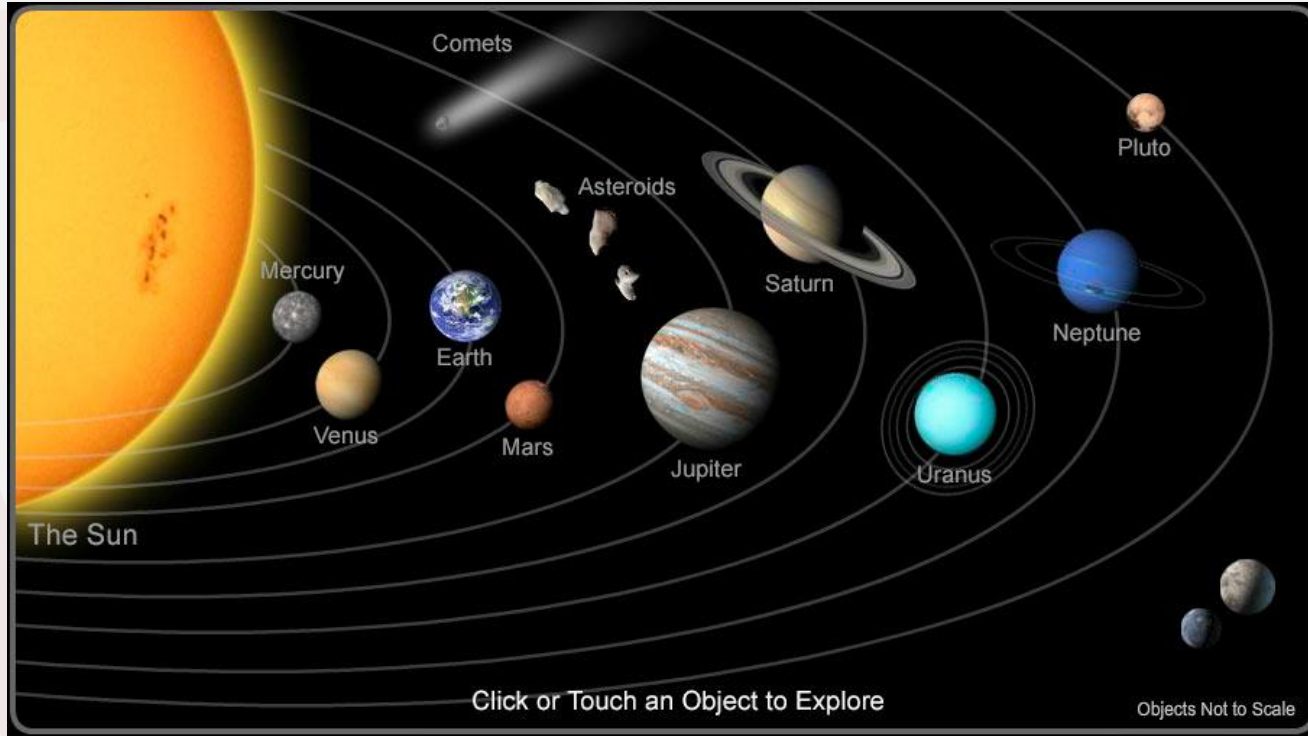
2016 1.000 USD



Entretenimento



Astrofísica / Astronomia



Telecomunicação
Varejo Internet Biologia
Saúde Manufatura Jogos
Finanças Energia
Governo Farmácia
Entretenimento



Faculdade de Computação e Informática
Mackenzie

Definição do projeto da disciplina



Referências (Materiais)

<http://www2.datasus.gov.br> | <http://dados.gov.br> | <https://www.data.gov>

<http://opendata.cern.ch/search?ln=en&p=Run2011A+AND+collection%3ACMS-Primary-Datasets+OR+collection%3ACMS-Simulated-Datasets+OR+collection%3ACMS-Derived-Datasets>

<https://databricks.com/blog/2016/05/24/genome-sequencing-in-a-nutshell.html>



■ Referências (Imagens)

<http://dilbert.com/strip/2013-02-02>

<https://dilbert.com/strip/2017-09-07>

<http://pt.dreamstime.com/photos-images/esquema-da-equipe-de-futebol.html#details13993885>

<http://pt.dreamstime.com/imagem-de-stock-royalty-free-vetor-cor-de-rosa-do-unicrnio-image10564106>

<http://www.seasky.org/solar-system/solar-system.html>





Faculdade de Computação e Informática
Mackenzie