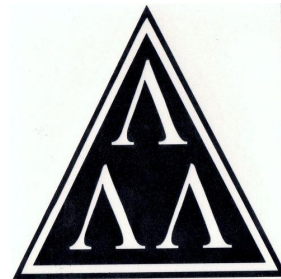


# O que ninguém conta sobre Data science

Eduardo Mendes  
[github.com/z4r4tu5tr4](https://github.com/z4r4tu5tr4)

babbage@z4r4tu5tr4: screenfetch



**Nome:** Eduardo Mendes

**Instituição:** Fatec Americana

**Uptime:** 12097080s

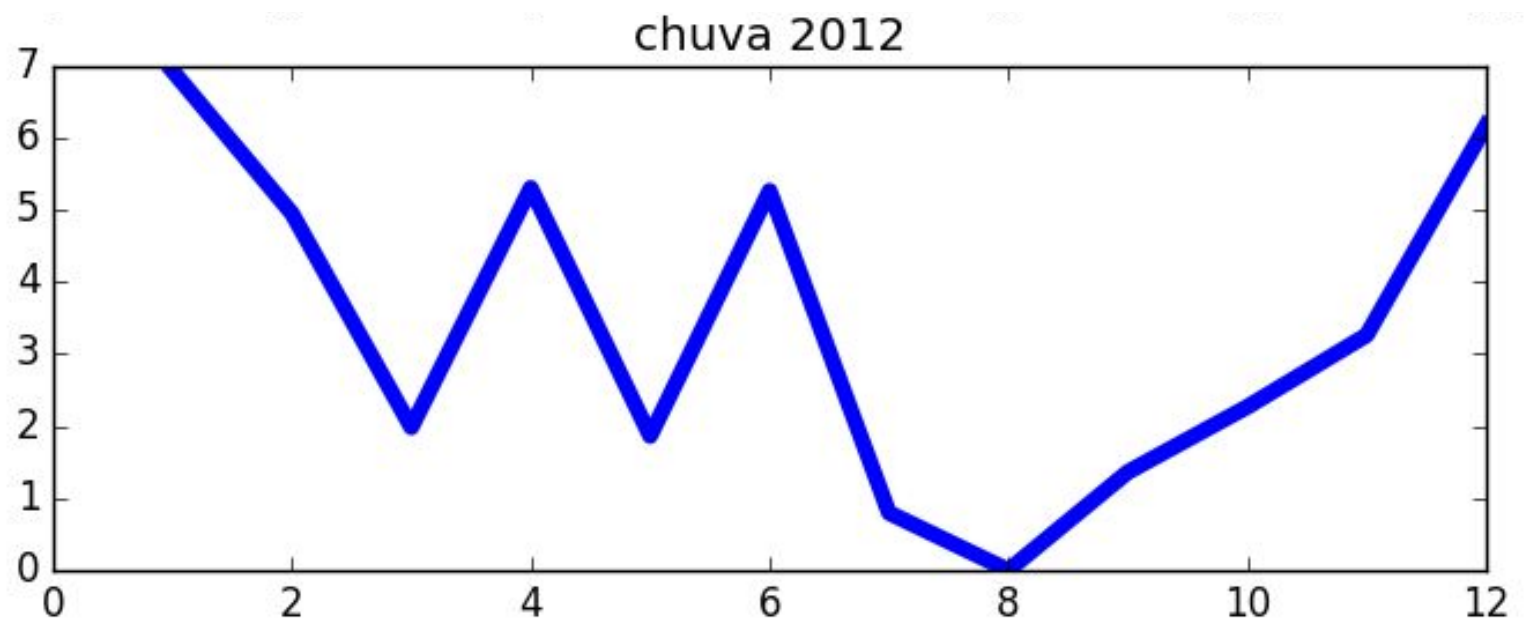
**Email:** mendesxeduardo@gmail.com

**git:** github.com/z4r4tu5tr4

## Nem sempre tudo é lindo como nos exemplos

**Esse  
mês não  
choveu**

201201	6.932258064516128
201202	4.953571428571428
201203	1.9838709677419355
201204	5.306666666666667
201205	1.864516129032258
201206	5.266666666666666
201207	0.7967741935483871
201208	0.0
201209	1.3633333333333335
201210	2.267741935483871
201211	3.263333333333333
201212	6.174193548387096



A gente ignora  
isso e bola pra  
frente



MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO  
INSTITUTO NACIONAL DE METEOROLOGIA - INMET  
ESTAÇÃO METEOROLÓGICA AUTOMÁTICA DE PIRACICABA/SP

Alt. 573m  
Lat. 22°42'S  
Lon. 47°37'W

		PRESSÃO ATMOSFERICA (hPa)	PRESSÃO ATMOSFERICA (hPa)	PRESSÃO ATMOSFERICA (hPa)	PRESSÃO ATMOSFERICA (hPa)	PRESSÃO ATMOSFERICA (hPa)	PRESSÃO ATMOSFERICA (hPa)	PRESSÃO ATMOSFERICA (hPa)	PRESSÃO ATMOSFERICA (hPa)
11	HORA UTC	0000	0100	0200	0300	0400	0500	0600	0700
12	01-JAN-2009	949,8	947,9	948,0	948,8	948,2	947,8	948,0	948,5
13	02-JAN-2009	946,9	947,3	947,4	947,1	946,8	946,6	946,0	945,7
14	03-JAN-2009	944,4	944,6	944,9	944,3	943,4	942,6	942,4	942,2
15	04-JAN-2009	941,2	941,6	941,7	941,9	941,7	941,8	941,9	942,0
16	05-JAN-2009	946,5	947,3	947,3	947,4	947,0	946,4	946,3	946,2
17	06-JAN-2009	947,8	948,4	948,7	948,4	948,2	947,9	947,5	947,3
18	07-JAN-2009	948,9	949,6	949,3	949,0	948,5	948,1	947,9	948,2
19	08-JAN-2009	948,1	948,8	948,9	948,6	948,5	948,1	948,0	948,0
20	09-JAN-2009	949,1	950,0	950,3	950,4	950,2	950,0	949,7	949,9
21	10-JAN-2009	952,5	954,0	954,1	953,6	953,2	952,9	952,8	952,8
22	11-JAN-2009	951,5	952,0	952,0	951,5	951,4	951,1	951,3	951,1

	V	W	X	Y	Z				AD
3									
4									
5									
6									
7									
8									
9									
10	PRESSÃO ATMOSFERICA (hPa)	PRESSÃO ATMOSFERICA (hPa)	PRESSÃO ATMOSFERICA (hPa)	PRESSÃO ATMOSFERICA (hPa)	PRESSÃO ATMOSFÉRICA MÁXIMA (hPa)	PRESSÃO ATMOSFERICA MÁXIMA (hPa)	PRESSÃO ATMOSFÉRICA MÁXIMA (hPa)	PRESSÃO ATMOSFÉRICA MÁXIMA (hPa)	PRESSÃO ATMOSFÉRICA MÁXIMA (hPa)
11	2000	2100	2200	2300	0000	0100	0200	0300	0400
12	946,2	946,0	946,2	946,6	949,8	950,0	948,2	948,9	948,8
13	943,5	943,0	943,2	943,8	947,1	947,3	947,5	947,4	947,1
14	940,2	940,2	940,7	941,2	944,4	944,6	945,0	944,9	944,3
15	944,3	944,6	945,1	945,6	941,5	941,6	941,7	941,9	941,9
16	945,1	945,6	946,4	947,2	946,5	947,3	947,6	947,5	947,4
17	946,3	946,9	946,9	948,1	947,9	948,4	948,7	948,7	948,4
18	946,3	946,5	947,0	947,4	948,9	949,6	949,6	949,3	949,0
19	947,3	947,5	947,9	948,2	948,1	948,8	949,0	949,0	948,6
20	948,2	950,6	950,9	951,8	949,1	950,0	950,3	950,4	950,6
21	950,0	950,5	950,8	950,9	952,5	954,0	954,2	954,3	953,6
22	949,8	950,0	951,1	951,3	951,5	952,0	952,2	952,0	951,5

Oi??

	HA	HB	HC	HD	HE	HF	HG	HH	HI
4									
5									
6									
7									
8									
9									
10	UMIDADE RELATIVA DO AR (%)	UMIDADE RELATIVA DO AR (%)	UMIDADE RELATIVA DO AR (%)	UMIDADE RELATIVA DO AR (%)	UMIDADE RELATIVA DO AR (%)	UMIDADE RELATIVA DO AR (%)	UMIDADE RELATIVA DO AR (%)	UMIDADE RELATIVA DO AR (%)	UMIDADE RELATIVA DO AR (%)
11	1500	1600	1700	1800	1900	2000	2100	2200	2300
12	68	69	62	62	61	60	65	71	74
13	79	79	78	75	75	74	79	82	83
14	71	83	80	80	81	72	78	81	83
15	84	93	93	85	77	70	73	76	80
16	65	61	57	59	58	56	62	71	75
17	52	54	47	52	76	58	63	62	66
18	54	40	34	51	49	40	35	49	65
19	39	36	37	37	35	32	32	35	42
20	42	47	39	36	34	42	83	89	89
21	55	57	52	49	54	59	71	77	81
22	57	53	49	49	47	51	63	78	83

```
In [1]: from string import ascii_uppercase
```

```
In [2]: ascii_uppercase.index("H")
```

```
Out[2]: 7
```

```
In [3]: 26*7*365*100
```

```
Out[3]: 6643000
```



```
In [1]: from string
```

```
In [2]: ascii_upperc
```

```
Out[2]: 7
```

```
In [3]: 26*7*365*100
```

```
Out[3]: 6643000
```

```
time ./met.py
```

```
real      355m31.871s
```

```
user      303m1.318s
```

```
sys       5m4.070s
```

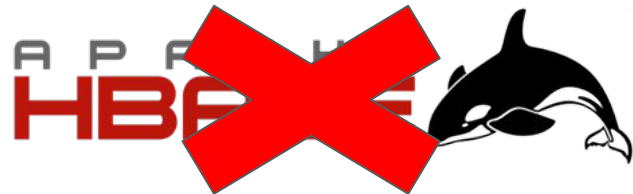
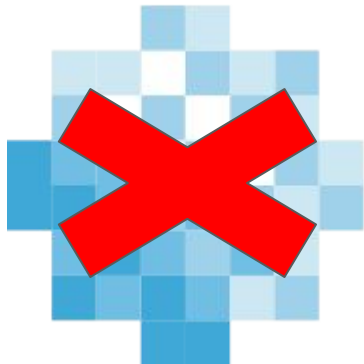
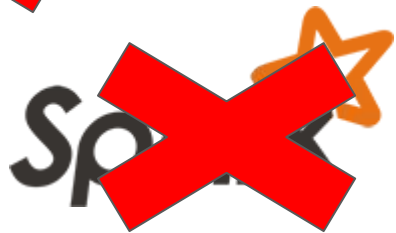
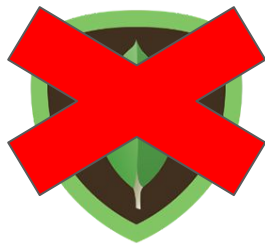
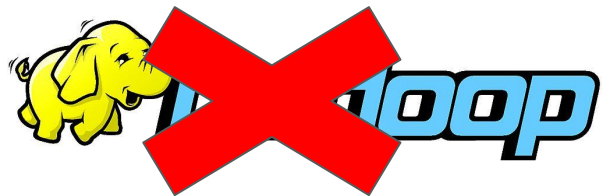
# **Se a quantidade for de mais, quem paga o preço?**

- Programadores?
- Infraestrutura?
- O Papa?

# Um pouco sobre processamento de massas [0]

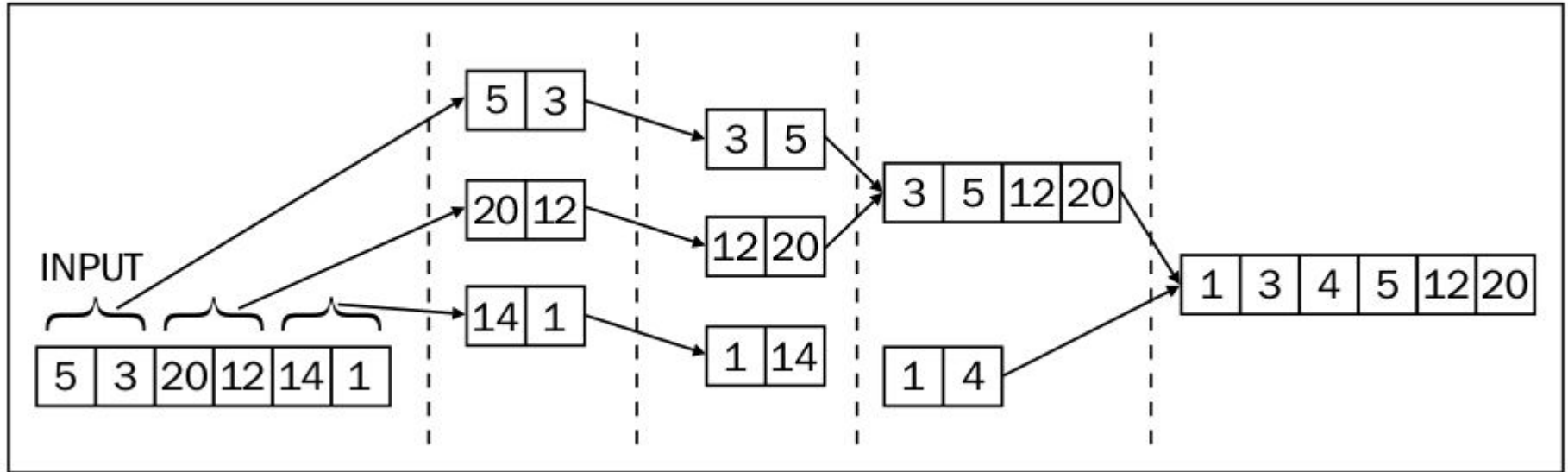


# Um pouco sobre processamento de massas [1]



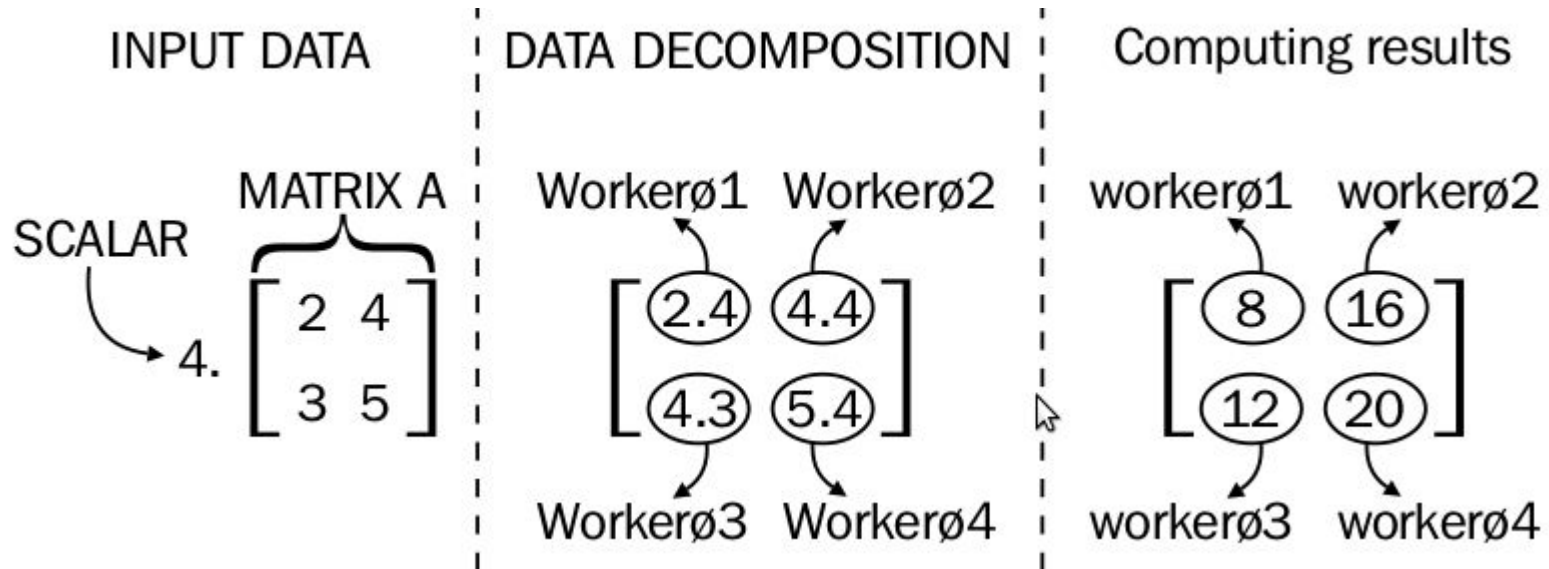
# Um pouco sobre processamento de massas [2]

- Dividir e conquistar



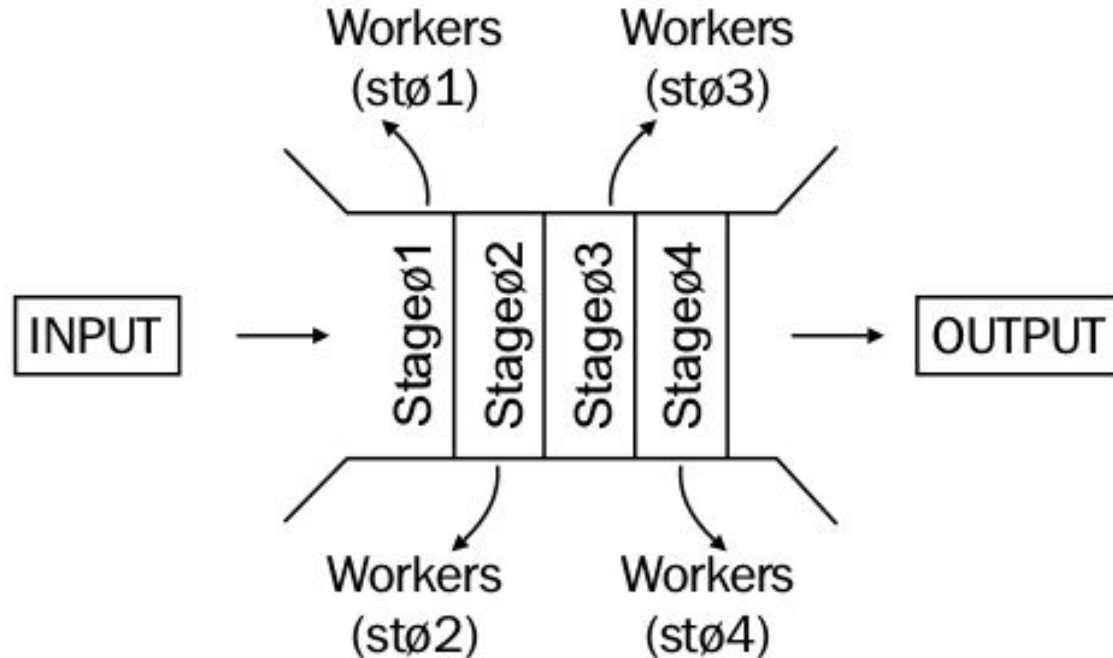
# Um pouco sobre processamento de massas [3]

- Decomposição de dados



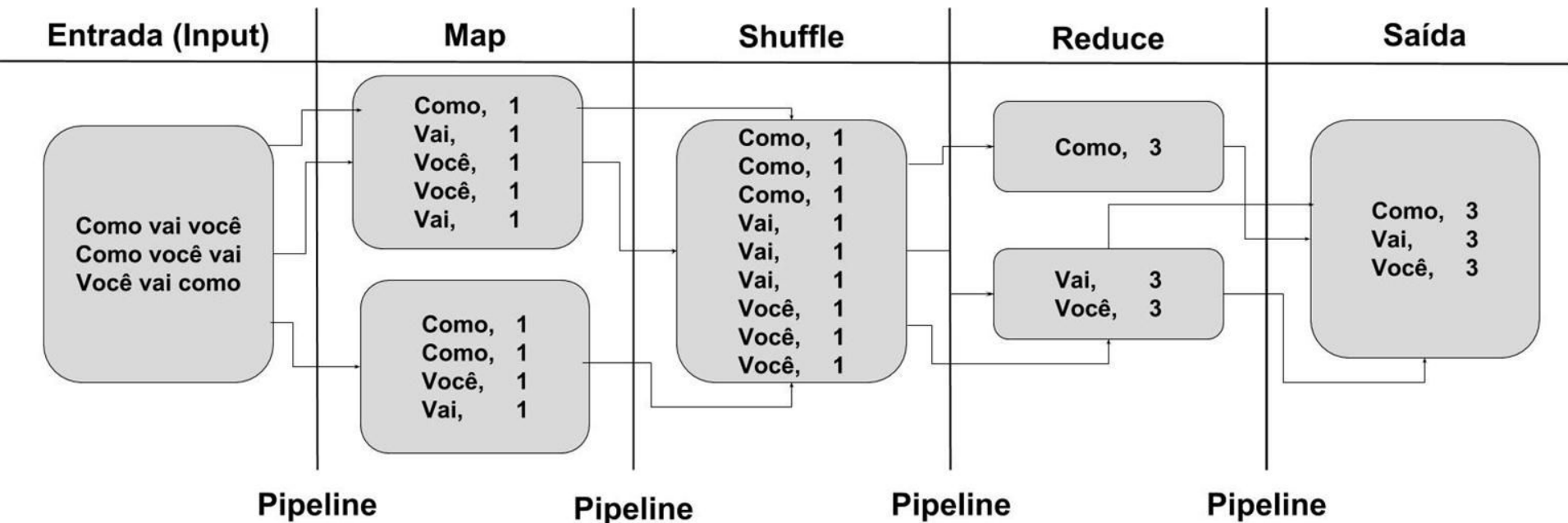
# Um pouco sobre processamento de massas [4]

- Pipeline



# Um pouco sobre processamento de massas [5]

- MapReduce



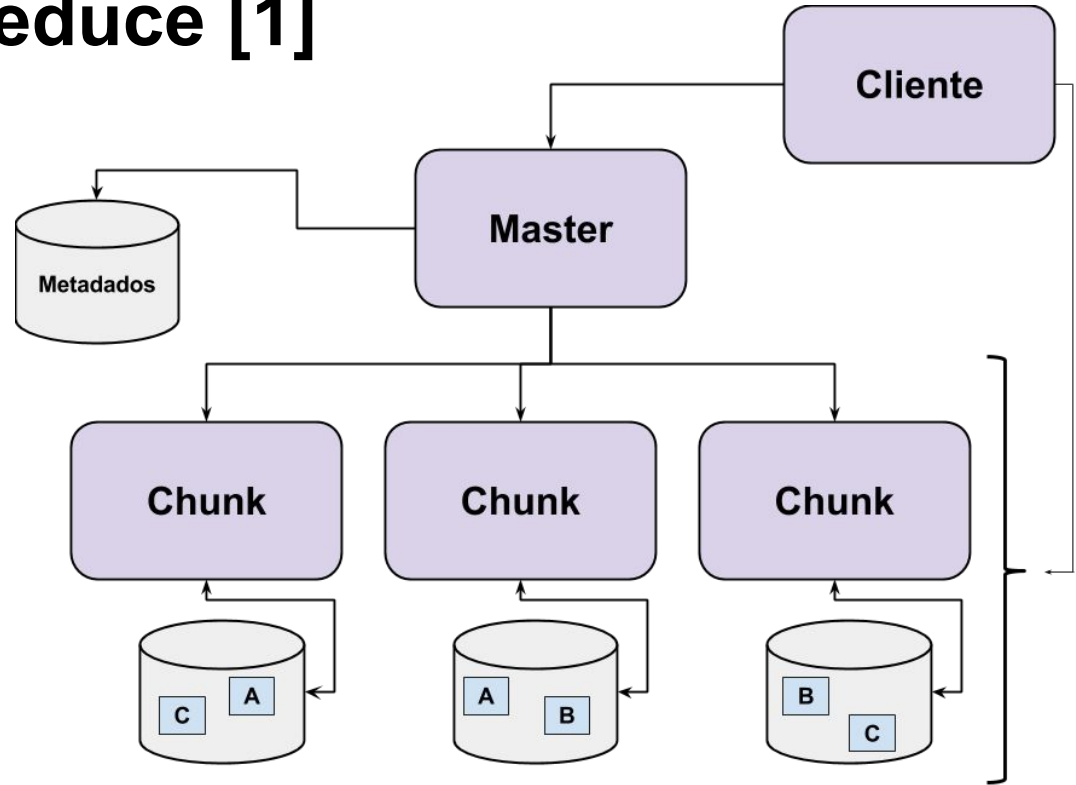
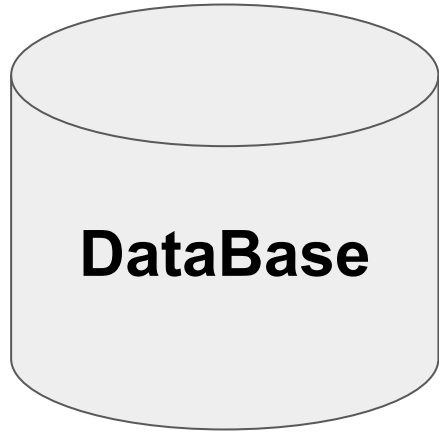


Ok! O processamento  
agora parece menos  
mediocre, **mas ...**

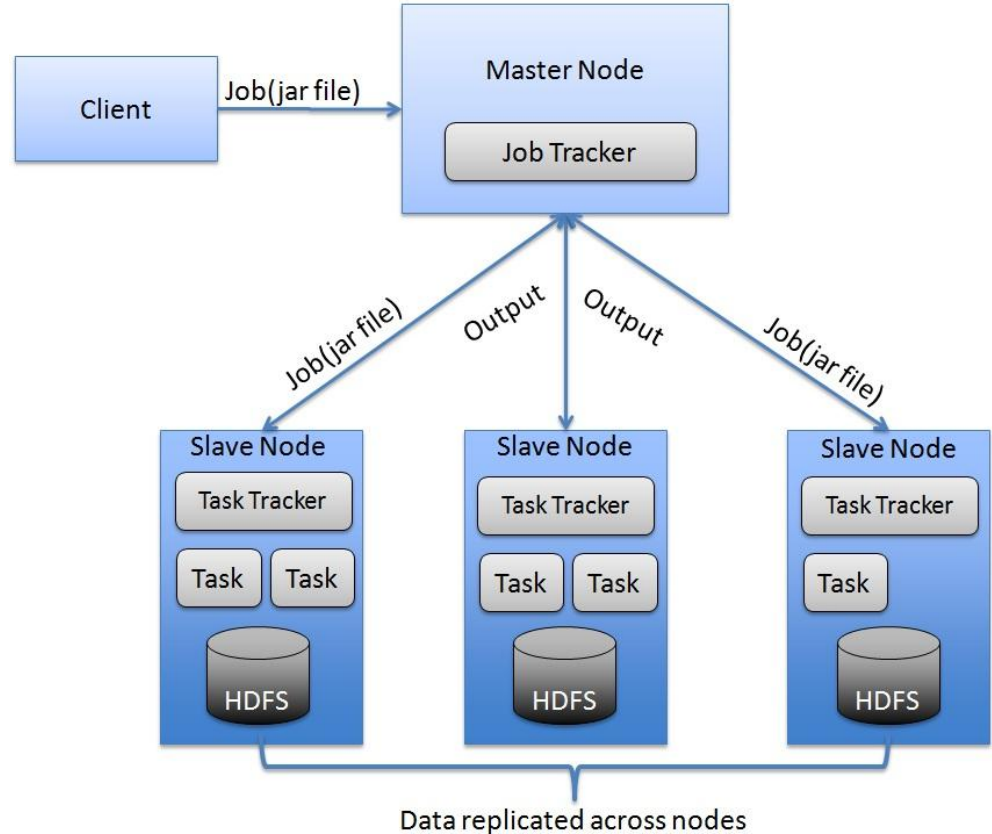
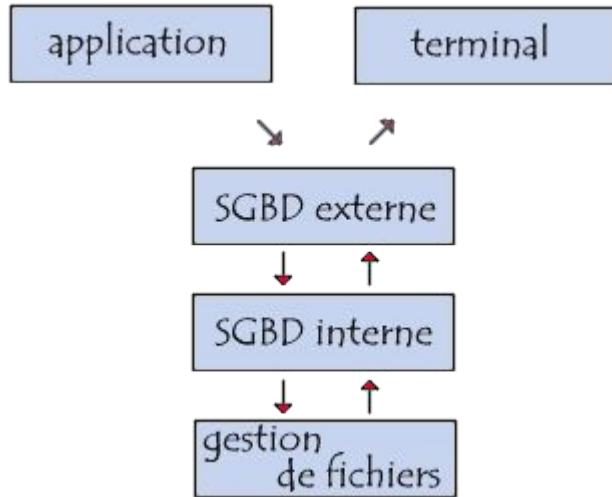
# DATABASE x MapReduce [0]

	Traditional RDBMS	MapReduce
<b>Data size</b>	Gigabytes	Petabytes
<b>Access</b>	Interactive and batch	Batch
<b>Updates</b>	Read and write many times	Write once, read many times
<b>Transactions</b>	ACID	<b>None</b>
<b>Structure</b>	Schema-on-write	Schema-on-read
<b>Integrity</b>	High	Low
<b>Scaling</b>	Nonlinear	Linear

# DATABASE x MapReduce [1]

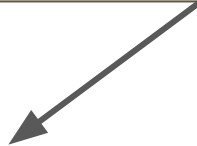


# DATABASE x MapReduce [2]



# DATABASE x MapReduce [3]

**“O rato roeu a roupa do rei de roma”**



E aquela análise  
meteorológica feita  
distribuída, como fica?

1 nó	10 nós
355m 1.318s	97m 56.008s

# XOXO!

Dúvidas?

[mendesxeduardo@gmail.com](mailto:mendesxeduardo@gmail.com)



# Considerações

- **Figuras slide 13-15 (Parallel Programming with Python - Jan Palach)**
- **Tabela Slide 18 (Hadoop - The definitive guide - Tom White)**

**Me desculpem pelo gráfico do slide 4, fiz só pra provocar mesmo.**