

## Sumário

Introdução.....	1
Databricks - Lakeflow Jobs .....	1
Databricks - SQL Warehouse .....	2
Databricks - Simple Storage Service (S3) .....	3
S3 – custo total .....	3
S3 – armazenamento.....	4
S3 – transferência outbound.....	4
S3 – operações.....	4
Total: Databricks + AWS.....	5

## Introdução

Neste documento, orçamos o custo para utilizar o Databricks na AWS. Para isso consideramos três cenários de volumetria.

Ressalta-se que não foram estimados custos de infraestrutura para as soluções que venham a ser futuramente adotadas pela Costal nos domínios de gestão de obras, finanças e compras, recursos humanos, comercial e gestão de contratos, uma vez que tais soluções são, em sua maioria, disponibilizadas no modelo Software as a Service (SaaS).

Nesse modelo, a infraestrutura tecnológica necessária à sua operação — incluindo processamento, armazenamento, segurança e disponibilidade — é provida pelos próprios fornecedores e já se encontra incorporada ao preço dos serviços contratados.

## Databricks - Lakeflow Jobs

Calculadora de custos utilizada:

<https://www.databricks.com/product/pricing/product-pricing/instance-types>

### Tabela de custos

Cálculo - Lakeflow jobs						
Cenário	Instancias EC2	Horas/ dia	Dias/ mês	USD/ mês	USD/ mês (prd + hml)	
1	10	10	8	30	270	540
2	20	20	10	30	675	1350
3	30	30	12	30	1215	2430

Consideramos três cenários com diferentes volumetrias.

Teremos dois ambientes: um de produção e um para testes. Por isso, o custo total foi multiplicado por 2.

### Parâmetros considerados

Esses foram os parâmetros definidos na calculadora do Databricks para estimar o custo. O valor cobrado em dólares por hora depende dessas especificações.

Parâmetros - Lakeflow jobs				
Plano Databricks	Compute type	Instance type	Região	USD/ hora
Premium	Lakeflow jobs classic	m5.xlarge (4vCPUs, 16GB)	N. Virgínia	0.1125

### Informações adicionais

O custo por hora já considera a infraestrutura AWS + o DBU do Databricks.

O Lakeflow Jobs Classic usa instâncias EC2 para executar os jobs. Os servidores são desligados após a execução dos jobs.

Existem componentes opcionais do Databricks que não foram considerados nessa estimativa. Se decidirmos usá-los no futuro, podemos incluir:

- Machine Learning
- Interactive Workloads (Notebooks)

O cluster deve ter no mínimo 2 servidores: 1 driver e 1 worker. O driver gerencia os workers.

A quantidade de servidores foi definida considerando também tarefas auxiliares, como:

- Tarefas de manutenção do Databricks (OPTIMIZE, VACUUM, COMPUTE STATISTICS).
- Tarefas de otimização automática (ex.: cluster on write).
- Cálculo, armazenamento e manutenção de metadados.

## Databricks - SQL Warehouse

Calculadora de custos utilizada:

<https://www.databricks.com/product/pricing/product-pricing/instance-types>

### Tabela de custos

Cálculo - SQL Warehouse				
Cenário	Instances	USD / mês	USD / mês (prd + hml)	
1	2	7603.2	15206.4	
2	3	11404.8	22809.6	
3	4	15206.4	30412.8	

Consideramos três cenários com diferentes volumetrias.

Teremos dois ambientes: um de produção e um para testes. Por isso, o custo total foi multiplicado por 2.

### Parâmetros utilizados

Esses foram os parâmetros definidos na calculadora do Databricks para estimar o custo. O valor cobrado em dólares por hora depende dessas especificações.

Parâmetros - SQL Warehouse					
Plano Databricks	Compute type	Compute configuration	Horas / dia	Dias / mês	USD / hora
Premium	SQL Compute	Medium	24	30	5.28

O front-end (cliente) fará consultas sobre os dados no Databricks, e isso exigirá um SQL Warehouse ligado 24h por dia nos três cenários.

Consideramos os ambientes de produção e o de testes ligados 24h por dia.

## Databricks - Simple Storage Service (S3)

Calculadora de custos utilizada: <https://calculator.aws/>

O Databricks executando na AWS utiliza o S3 para armazenamento.

### S3 – custo total

#### Tabela de custos

Cálculo - S3 - total	
Cenário	USD / mês
1	112
2	195
3	278

Consideramos três cenários com diferentes volumetrias.

Este é o custo total. A seguir detalharemos os três serviços associados ao S3, que somados resultam no custo total.

- Armazenamento
- Transferência
- Operações (PUT, COPY, POST LIST, GET, SELECT)

### Parâmetros utilizados

Esses foram os parâmetros definidos na calculadora da AWS para estimar o custo. Os valores cobrados dependem dessas especificações.

Parâmetros - S3 - configurações	
Região	Classe de armazenamento
N. Virgínia	Standard

## S3 – armazenamento

### Tabela de custos

Cálculo - S3 - armazenamento			
Cenário	GB/ mês	USD/ mês	
1	2000	40	
2	3000	60	
3	4000	80	

Consideramos três cenários com diferentes volumetrias.

### Parâmetros utilizados

Esse é o valor cobrado em dólares por GB considerando a região e a classe de armazenamento escolhidas.

Parâmetros - S3 - armazenamento	
Serviço	USD/ GB
Armazenamento	0.02

## S3 – transferência outbound

### Tabela de custos

Cálculo - S3 - transferência outbound			
Cenário	GB/ mês	USD/ mês	
1	200	18	
2	300	27	
3	400	36	

Consideramos três cenários com diferentes volumetrias.

### Parâmetros utilizados

Esse é o valor cobrado em dólares por GB considerando a região e a classe de armazenamento escolhidas.

Parâmetros - S3 - transferência outbound	
Serviço	USD/ GB
Transferência	0.09

## S3 – operações

### Tabela de custos

Cálculo - S3 - operações			
Cenário	PUT, COPY, POST, LIST / mês	GET, SELECT / mês	USD / mês
1	10000000	10000000	54
2	20000000	20000000	108
3	30000000	30000000	162

Consideramos três cenários com diferentes volumetrias.

### Parâmetros utilizados

Esse é o valor cobrado por requisição considerando a região e a classe de armazenamento escolhidas.

Parâmetros - S3 - operações	
Operação	USD / requisição
GET, SELECT	0.0000004
PUT, COPY, POST, LIST	0.000005

### Informações adicionais

Além do armazenamento, a AWS cobra a transferência de dados para fora da região em que os dados estão armazenados (outbound). Por exemplo: a transferência de dados da N. Virginia para a Internet ou outras regiões é cobrada.

A transferência inbound (internet para AWS) não é cobrada.

Há ainda a cobrança por requisições GET, SELECT e PUT, COPY, POST, LIST.

Para estimar o volume e a frequência de atualização dos dados consideramos o ambiente de produção e desenvolvimento.

## Total: Databricks + AWS

Tabela mostrando o custo total, considerando os Lakeflow Jobs, S3 e SQL Warehouse.

Cálculo - total - Databricks na AWS				
Cenário	Lakeflow jobs   USD / mês	S3   USD / mês	SQLWarehousing   USD / mês	Total   USD / mês
1	540	112	15206.4	15858.4
2	1350	195	22809.6	24354.6
3	2430	278	30412.8	33120.8