



Data Master

Case de Eng. de Dados

MARCIO DE LIMA - JUN/22



Contexto de Negócio



A empresa RoboCar (locadora de veículos autônomos), localizada no estado de São Paulo-Capital, deseja implementar um Data Lake em seu ambiente on-premise.

O objetivo é coletar e armazenar historicamente seus dados de locação, veículos, clientes, etc.; para gerar análises históricas assim como transformar os dados em KPIs para tomadas de decisão.



Indicadores de desempenho - KPI



A empresa RoboCar deseja extrair de seu Data Lake os indicadores abaixo:

- Quais e quantos veículos foram locados no grão mensal;
- Quais despachantes locaram no grão mensal;
- Qual o faturamento por veículo e despachantes no grão mensal;
- Quais contratos são de tipo de Risco (Aluguel para menores de 24 anos) com grão mensal;
- Acompanhar on-line a entrada de novos contratos durante o dia;



Dados

/datalake



raw



bronze



silver



gold



/tabela/data/arquivos

/tabela/data/arquivos

/tabela/arquivos

/tabela/arquivos

Dados brutos ingeridos de várias fontes (dados RDBMS, arquivos JSON, dados IoT etc.).

Parquet / JSON

Dados brutos transformados em tabela colunar e dados mascarados, conforme LGPD.

Parquet

Aplicação de Data Quality, de-duplicação de dados, tratamentos de tipagem de colunas, filtros e transformações.

Parquet

Dados com agregações de nível de negócios frequentemente usadas para painéis e relatórios.

Parquet



Acesso aos dados – Hdfs ACLs

As listas de controle de acesso podem ser usadas para implementar uma ampla variedade de controles de segurança de forma eficaz, mas depende muito do caso de uso, como sabemos, não existe “bala de prata”, cada cenário deve ser analisado para verificar se o ACLs é a melhor solução.

Vantagens

- Simplicidade => Uma ACL define claramente os níveis de acesso e permissões que cada usuário, grupo ou dispositivo possui em um determinado sistema. Isso facilita a definição e a interpretação de uma ACL. Como essas listas podem ser facilmente legíveis por humanos, um administrador pode determinar facilmente as permissões atuais e os controles de acesso colocados em um sistema, fazer edições e revogar permissões conforme necessário.

Desvantagens

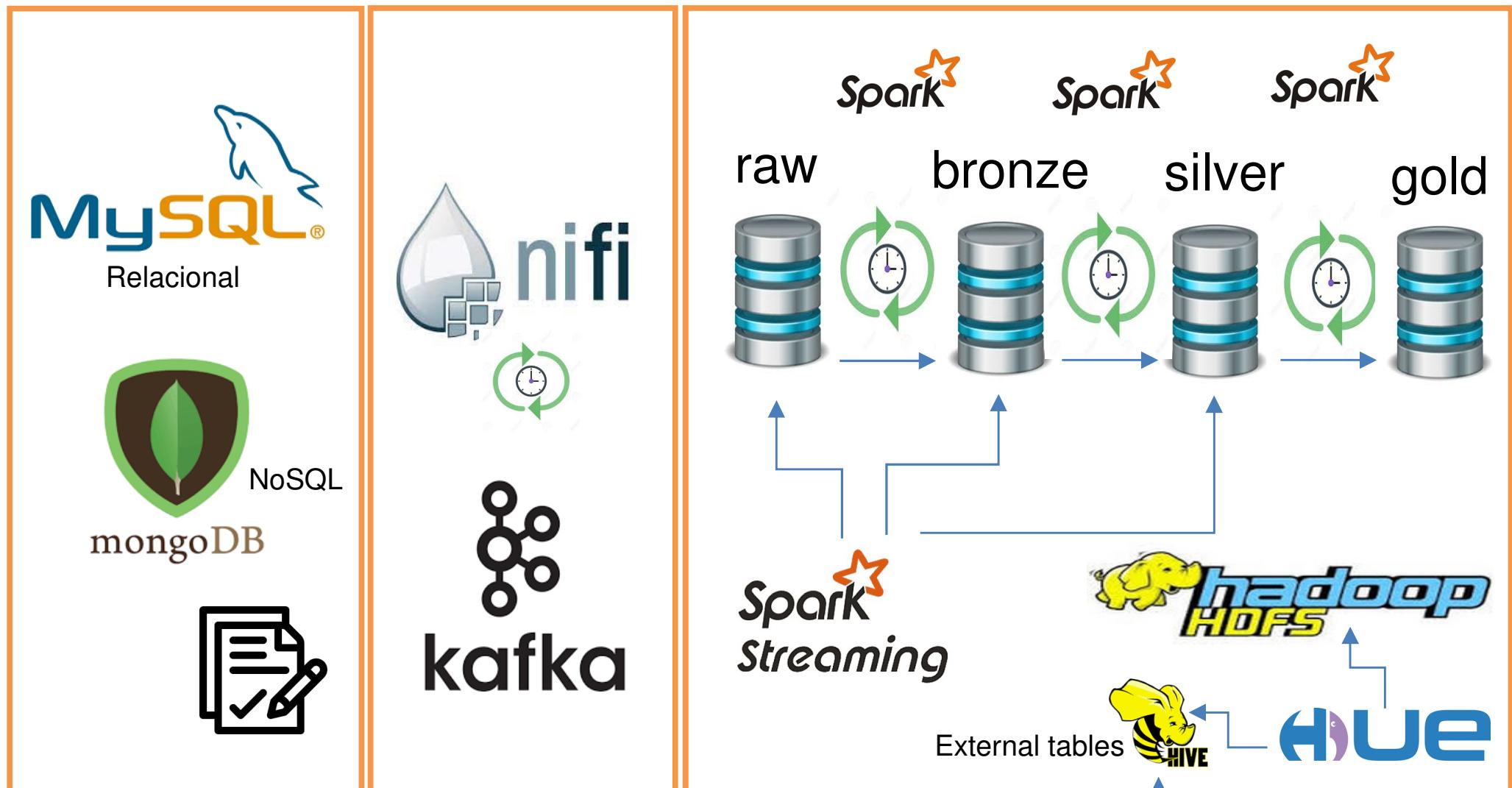
- Eficiência => As ACLs carecem de eficiência, pois suportam apenas controles de acesso explicitamente declarados. Se, por exemplo, um usuário tiver acesso ou permissões exclusivas porque pertence ao departamento de TI e a um gerente, esse nível de acesso deve ser declarado explicitamente em vez de inferido com base na participação em ambos os grupos;
- Escalabilidade => A medida que o número de usuários, grupos e recursos aumenta, aumenta também o tamanho da ACL e o tempo necessário para determinar o nível de acesso concedido a um determinado usuário. Além disso, quanto maior a lista maior o risco de falha humana do administrador nas liberações de acessos;
- Visibilidade => As permissões e os níveis de acesso de um usuário / grupo podem estar espalhados por várias listas independentes. Auditoria, alteração ou revogação de acesso requer uma revisão de cada ACL no ambiente da organização para aplicar as novas permissões.

Desenho da solução

Dados de origem

Ingestão

DataLake



Servidores – VM's GCP Cloud

console.cloud.google.com/compute/instances?hl=pt_BR&project=engdados-santander

Status do período de teste gratuito: R\$ 411,01 de crédito e 74 dias restantes. Com uma conta completa, você tem acesso ilimitado a todos os recursos do Google Cloud Platform.

DISPENSAR ATIVAR

Google Cloud Platform engdados-santander Pesquisa Produtos, recursos, documentos (/)

Compute Engine Instâncias da ... CRIAR INSTÂNCIA OPERAÇÕES MOSTRAR PAINEL DE INFORMAÇÕES SAIBA MAIS

<input type="checkbox"/> Status	Nome	Zona	Recomendações	Em uso por	IP interno	IP ext.	Conectar
<input checked="" type="checkbox"/>	datanode	southamerica-east1-a			0.219	34.95 (nic0)	SSH
<input checked="" type="checkbox"/>	master	southamerica-east1-a			10.10.0.220 (nic0)	35.24 (nic0)	SSH
<input checked="" type="checkbox"/>	namenode	southamerica-east1-a					
<input checked="" type="checkbox"/>	tools1	southamerica-east1-b					
<input checked="" type="checkbox"/>	worker	southamerica-east1-a					

 **CentOS** 



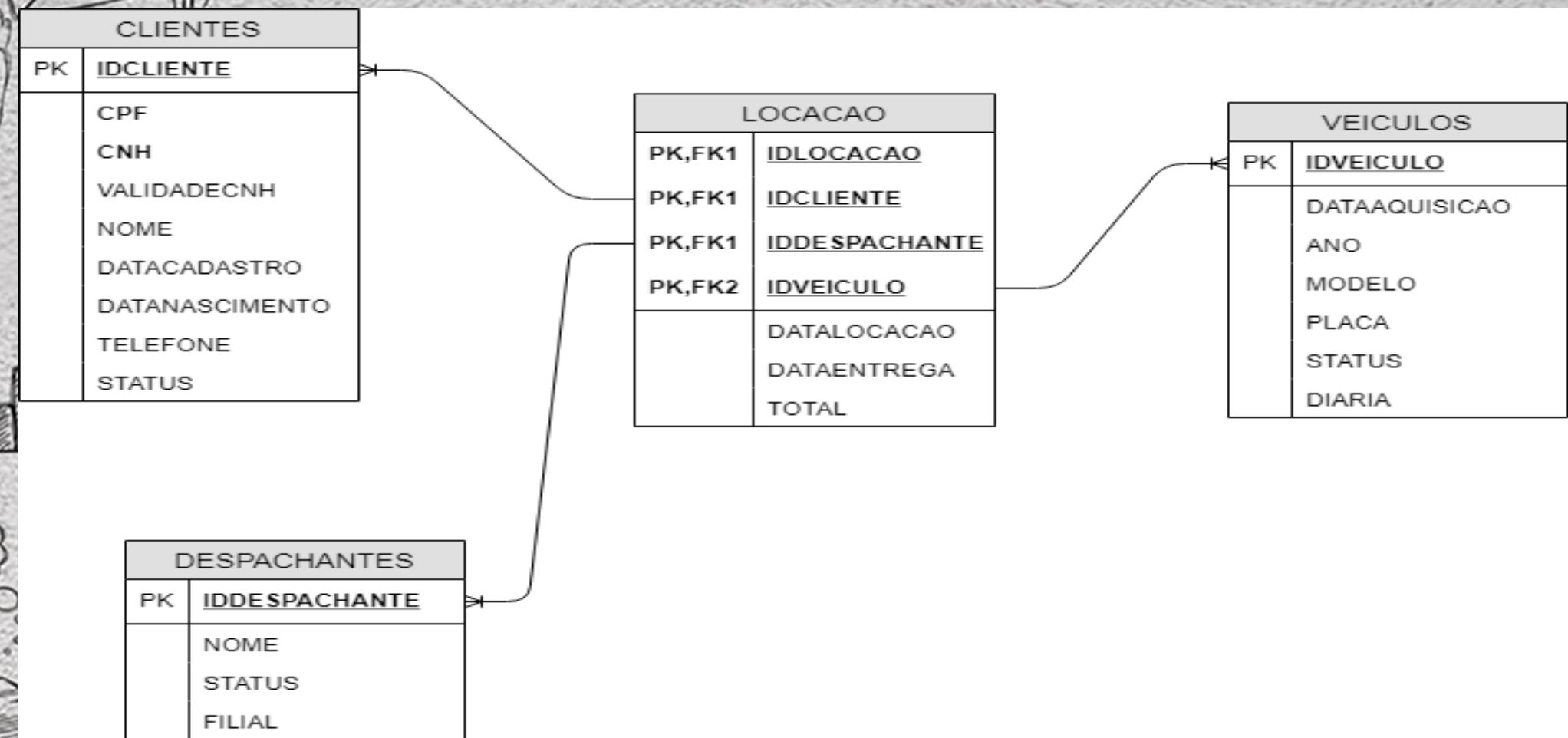







Marketplace Notas de lançamento

Dados - Origens



```

{
    "_id" : ObjectId("5d3b13edd6b1f54a1234108a"),
    "idcliente" : "2",
    "idcontrato" : "22082",
    "datacadastro" : "15/01/2019",
    "contrato" : "CONTRATO DE LOCAÇÃO DE AUTOMÓVEL DE PRAZO DETERMINADO IDENTIFICAÇÃO DAS PARTES CONTRATANTES LOCADORA: RoboCar Tipo de Contrato: de risco  

    to e acertado o presente Contrato de Locação de Automóvel de Prazo Determinado, que se regerá pelas cláusulas seguintes e pelas condições descritas no presente  

    ação do automóvel de propriedade da LOCADORA. DA USO Cláusula 2º. O automóvel, objeto deste contrato, será utilizado exclusivamente pelo LOCATÁRIO, não sendo  

    o da multa prevista na Cláusula 7º. DA DEVOLUÇÃO Cláusula 3º. O LOCATÁRIO deverá devolver o automóvel à LOCADORA nas mesmas condições em que estava quando o rec  

    ízios causados. Cláusula 5º. Se o LOCATÁRIO não restituir o automóvel na data estipulada, deverá pagar, enquanto detiver em seu poder, o aluguel que a LOCADORA a  

    niente de caso fortuito. DA RESCISÃO Cláusula 6º. É assegurado às partes a rescisão do presente contrato a qualquer momento, desde que haja comunicação à outra  

    quer das cláusulas por parte dos contratantes ensejará a rescisão deste instrumento e o devido pagamento de multa, pela parte inadimplente no valor de R$ 5000,  

    s vias de igual teor, juntamente com 2 (duas) testemunhas."
}

```

Dados - Raw

```
>>> dadosCli.head(1)
[Row(IDCLIENTE=1, CPF='16528447080', CNH='44971318961', VALIDADECNH='1632960000000', NOME='Alvito Espargosa', D
ATACADASTRO='1548374400000', DATANASCIMENTO='101001600000', TELEFONE='51390834589', STATUS='Ativo')]
>>>
>>> dadosVei.head(1)
[Row(IDVEICULO=1, DATAAQUISICAO='1548374400000', ANO=2019, MODELO='Tesla Model S P90D', PLACA='LWJ2929', STATUS
='Disponivel', DIARIA='1800.00')]
>>>
>>> dadosLoc.head(1)
[Row(IDLOCACAO=1, IDCLIENTE=3, IDDESPACHANTE=5, IDVEICULO=22, DATALOCACAO='1550620800000', DATAENTREGA='1550620
800000', TOTAL='1996.00')]
>>>
>>> dadosDes.head(1)
[Row(IDDESPACHANTE=1, NOME='Carminda Pestana', STATUS='Ativo', FILIAL='Santa Maria')]
```

```
>>>
>>> dadosCli.dtypes
[('IDCLIENTE', 'int'), ('CPF', 'string'), ('CNH', 'string'), ('VALIDADECNH', 'string'), ('NOME', 'string'), ('D
ATACADASTRO', 'string'), ('DATANASCIMENTO', 'string'), ('TELEFONE', 'string'), ('STATUS', 'string')]
>>>
>>> dadosVei.dtypes
[('IDVEICULO', 'int'), ('DATAAQUISICAO', 'string'), ('ANO', 'int'), ('MODELO', 'string'), ('PLACA', 'string'),
('STATUS', 'string'), ('DIARIA', 'string')]
>>>
>>> dadosLoc.dtypes
[('IDLOCACAO', 'int'), ('IDCLIENTE', 'int'), ('IDDESPACHANTE', 'int'), ('IDVEICULO', 'int'), ('DATALOCACAO', 's
tring'), ('DATAENTREGA', 'string'), ('TOTAL', 'string')]
>>>
>>> dadosDes.dtypes
[('IDDESPACHANTE', 'int'), ('NOME', 'string'), ('STATUS', 'string'), ('FILIAL', 'string')]
>>>
```

Bronze

```
>>> dadosCli.head(1)
[Row(IDCLIENTE=1, CPF='d5b84d3fbf1b14b398501f6759d4b9f1', CNH='dc9ebec7d697d0a08aa54b82c386127c', VALIDADECNH='1632960000000', NOME='ANONIMIZADO', DATACADASTRO='1548374400000', DATANASCIMENTO='101001600000', TELEFONE=0, STATUS='Ativo')]
>>>
>>> dadosVei.head(1)
[Row(IDVEICULO=1, DATAAQUISICAO='1548374400000', ANO=2019, MODELO='Tesla Model S P90D', PLACA='LWJ2929', STATUS='Disponivel', DIARIA='1800.00')]
>>>
>>> dadosLoc.head(1)
[Row(IDLOCACAO=1, IDCLIENTE=3, IDDESPACHANTE=5, IDVEICULO=22, DATALOCACAO='1550620800000', DATAENTREGA='1550620800000', TOTAL='1996.00')]
>>>
>>> dadosDes.head(1)
[Row(IDDESPACHANTE=1, NOME='ANONIMIZADO', STATUS='Ativo', FILIAL='Santa Maria')]
>>>
>>> dadosContratos.head(1)
[Row(contrato='CONTRATO DE LOCAÇÃO DE AUTOMÓVEL DE PRAZO DETERMINADO IDENTIFICAÇÃO DAS PARTES CONTRATANTES LOCADORA: RoboCar Tipo de Contrato: normal LOCATÁRIO: (Nome do Locatário)As partes acima identificadas têm, entre si, justo e acertado o presente Contrato de Locação de Automóvel de Prazo Determinado, que se regerá pelas cláusulas seguintes e pelas condições descritas no presente. DO OBJETO DO CONTRATO Cláusula 1ª. O presente contrato tem como OBJETO a locação do automóvel de propriedade da LOCADORA. DO USO Cláusula 2ª. O automóvel, objeto deste contrato, será utilizado exclusivamente pelo LOCATÁRIO, não sendo permitido o seu uso por terceiros sob pena de rescisão contratual e o pagamento da multa prevista na Cláusula 7ª.DA DEVOLUÇÃO Cláusula 3ª. O LOCATÁRIO deverá devolver o automóvel à LOCADORA nas mesmas condições em que estava quando o recebeu, ou seja, em perfeitas condições de uso, respondendo pelos danos ou prejuízos causados.Cláusula 5ª. Se o LOCATÁRIO não restituir o automóvel na data estipulada, deverá pagar, enquanto detiver em seu poder, o aluguel que a LOCADORA arbitrar, e responderá pelo dano, que o automóvel venha a sofrer mesmo se proveniente de caso fortuito3.DA RESCISÃO Cláusula 6ª. É assegurado às partes a rescisão do presente contrato a qualquer momento, desde que haja comunicação à outra parte com antecedência mínima de 2 dias.Cláusula 7ª. O descumprimento de qualquer das cláusulas por parte dos contratantes ensejará a rescisão deste instrumento e o devido pagamento de multa, pela parte inadimplente no valor de R$ 5000, Por estarem assim justos e contratados, firmam o presente instrumento, em duas vias de igual teor, juntamente com 2 (duas) testemunhas.', datacadastro='09/02/2019', idcliente='17', idcontrato='94815')]
>>> ■
>>>
>>> dadosCli.dtypes
[('IDCLIENTE', 'int'), ('CPF', 'string'), ('CNH', 'string'), ('VALIDADECNH', 'string'), ('NOME', 'string'), ('DATACADASTRO', 'string'), ('DATANASCIMENTO', 'string'), ('TELEFONE', 'string'), ('STATUS', 'string')]
>>>
>>> dadosVei.dtypes
[('IDVEICULO', 'int'), ('DATAAQUISICAO', 'string'), ('ANO', 'int'), ('MODELO', 'string'), ('PLACA', 'string'), ('STATUS', 'string'), ('DIARIA', 'string')]
>>>
>>> dadosLoc.dtypes
[('IDLOCACAO', 'int'), ('IDCLIENTE', 'int'), ('IDDESPACHANTE', 'int'), ('IDVEICULO', 'int'), ('DATALOCACAO', 'string'), ('DATAENTREGA', 'string'), ('TOTAL', 'string')]
>>>
>>> dadosDes.dtypes
[('IDDESPACHANTE', 'int'), ('NOME', 'string'), ('STATUS', 'string'), ('FILIAL', 'string')]
>>>
```

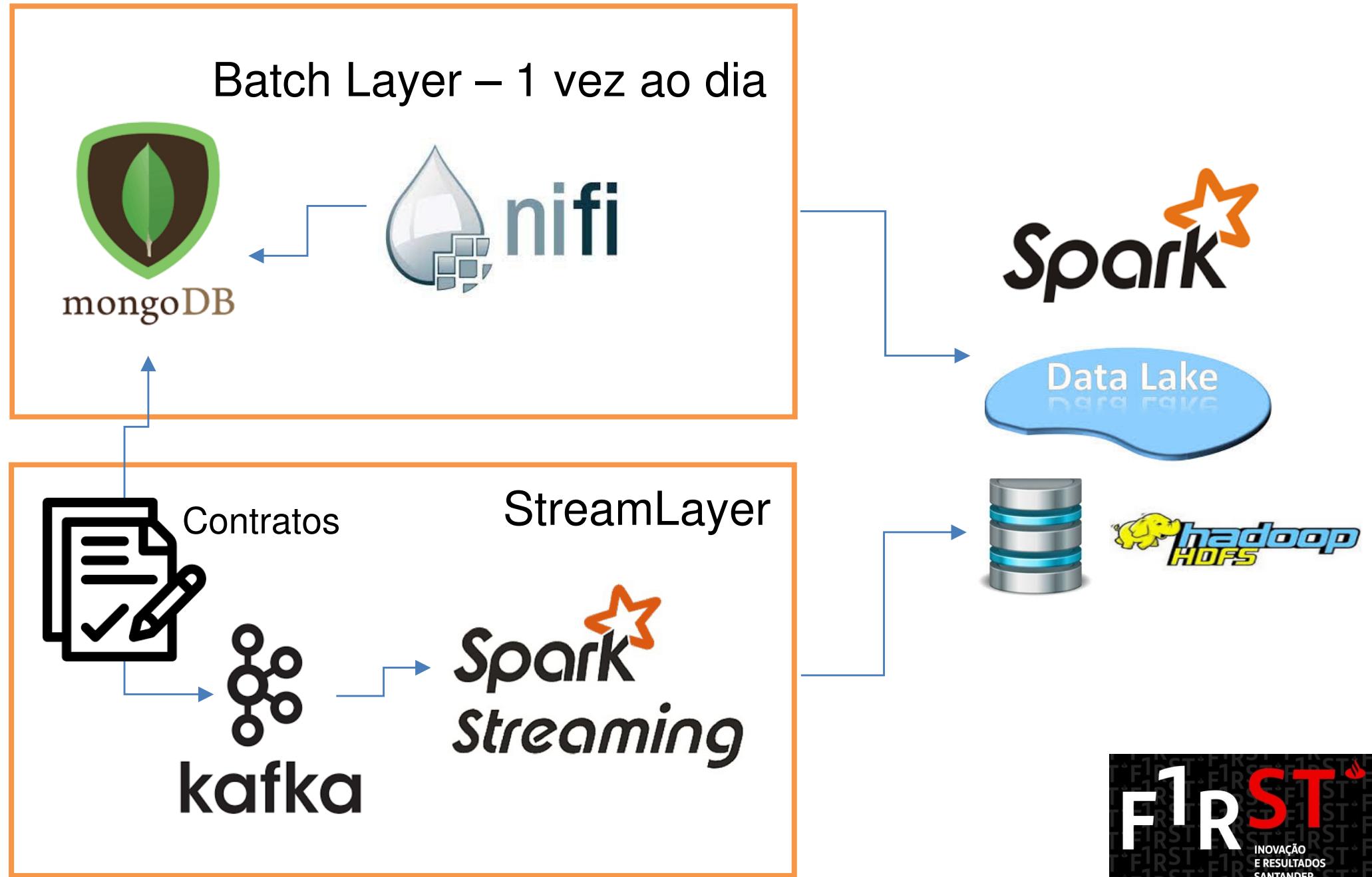
Silver

```
>>> dadosCli.head(1)
[Row(IDCLIENTE=1, CPF='d5b84d3fbf1b14b398501f6759d4b9f1', CNH='dc9ebec7d697d0a08aa54b82c386127c', VALIDADECNH='1632960000000', NOME='ANONIMIZADO', DATACADASTRO=datetime.date(2019, 1, 25), DATANASCIMENTO=datetime.date(1973, 3, 15), TELEFONE=0, STATUS='ATIVO')]
>>>
>>> dadosVei.head(1)
[Row(IDVEICULO=1, DATAAQUISICAO='1548374400000', ANO=2019, MODELO='TESLA MODEL S P90D', PLACA='LWJ2929', STATUS='DISPONÍVEL', DIARIA='1800.00')]
>>>
>>> dadosLoc.head(1)
[Row(IDLOCACAO=1, IDCLIENTE=3, IDDESPACHANTE=5, IDVEICULO=22, DATALOCACAO=datetime.date(2019, 2, 20), DATAENTREGA=datetime.date(2019, 2, 20), TOTAL='1996.00')]
>>>
>>> dadosDes.head(1)
[Row(IDDESPACHANTE=1, NOME='ANONIMIZADO', STATUS='ATIVO', FILIAL='SANTA MARIA')]
>>>
>>> dadosContratos.head(1)
[Row(CONTRATO='CONTRATO DE LOCAÇÃO DE AUTOMÓVEL DE PRAZO DETERMINADO IDENTIFICAÇÃO DAS PARTES CONTRATANTES LOCADORA: RoboCar Tipo de Contrato: normal LOCATÁRIO: (Nome do Locatário)As partes acima identificadas têm, entre si, justo e acertado o presente Contrato de Locação de Automóvel de Prazo Determinado, que se regerá pelas cláusulas seguintes e pelas condições descritas no presente. DO OBJETO DO CONTRATO Cláusula 1ª. O presente contrato tem como OBJETO a locação do automóvel de propriedade da LOCADORA. DO USO Cláusula 2ª. O automóvel, objeto deste contrato, será utilizado exclusivamente pelo LOCATÁRIO, não sendo permitido o seu uso por terceiros sob pena de rescisão contratual e o pagamento da multa prevista na Cláusula 7ª.DA DEVOLUÇÃO Cláusula 3ª. O LOCATÁRIO deverá devolver o automóvel à LOCADORA nas mesmas condições em que estava quando o recebeu, ou seja, em perfeitas condições de uso, respondendo pelos danos ou prejuízos causados.Cláusula 5ª. Se o LOCATÁRIO não restituir o automóvel na data estipulada, deverá pagar, enquanto detiver em seu poder, o aluguel que a LOCADORA arbitrar, e responderá pelo dano, que o automóvel venha a sofrer mesmo se proveniente de caso fortuito3.DA RESCISÃO Cláusula 6ª. É assegurado às partes a rescisão do presente contrato a qualquer momento, desde que haja comunicação à outra parte com antecedência mínima de 2 dias.Cláusula 7ª. O descumprimento de qualquer das cláusulas por parte dos contratantes ensejará a rescisão deste instrumento e o devido pagamento de multa, pela parte inadimplente no valor de R$ 5000, Por estarem assim justos e contratados, firmam o presente instrumento, em duas vias de igual teor, juntamente com 2 (duas) testemunhas.', DATACADASTRO=datetime.date(2019, 2, 9), IDCLIENTE='17', IDCONTRATO='94815')]
>>>
>>> dadosCli.dtypes
[(('IDCLIENTE', 'int'), ('CPF', 'string'), ('CNH', 'string'), ('VALIDADECNH', 'string'), ('NOME', 'string'), ('DATACADASTRO', 'date'), ('DATANASCIMENTO', 'date'), ('TELEFONE', 'int'), ('STATUS', 'string'))]
>>>
>>> dadosVei.dtypes
[(('IDVEICULO', 'int'), ('DATAAQUISICAO', 'string'), ('ANO', 'int'), ('MODELO', 'string'), ('PLACA', 'string'), ('STATUS', 'string'), ('DIARIA', 'string'))]
>>>
>>> dadosLoc.dtypes
[(('IDLOCACAO', 'int'), ('IDCLIENTE', 'int'), ('IDDESPACHANTE', 'int'), ('IDVEICULO', 'int'), ('DATALOCACAO', 'date'), ('DATAENTREGA', 'date'), ('TOTAL', 'string'))]
>>>
>>> dadosDes.dtypes
[(('IDDESPACHANTE', 'int'), ('NOME', 'string'), ('STATUS', 'string'), ('FILIAL', 'string'))]
>>>
>>> dadosContratos.dtypes
[(('CONTRATO', 'string'), ('DATACADASTRO', 'date'), ('IDCLIENTE', 'string'), ('IDCONTRATO', 'string'))]
```

Gold

```
>>>
>>> dadosLocacaoCli.head(1)
[Row(ID=1, MES=3, ANO=2019, TOTAL=3)]
>>>
>>> dadosLocacaoVei.head(1)
[Row(MODELO='BMW 750I XDRIVE', ANO=2019, MES=5, TOTAL=10)]
>>>
>>> dadosLocacaoDes.head(1)
[Row(ID=3, MODELO='BMW 750I XDRIVE', ANO=2019, MES=5, TOTAL=10)]
>>>
>>> dadosFaturamento.head(1)
[Row(ANO=2019, MES=2, TOTAL=23852.0)]
>>>
>>> dadosRisco.head(1)
[Row(ID='93963', IDCLIENTE='14')]
>>>
>>> ■
>>> dadosLocacaoCli.dtypes
[('ID', 'int'), ('MES', 'int'), ('ANO', 'int'), ('TOTAL', 'bigint')]
>>>
>>> dadosLocacaoVei.dtypes
[('MODELO', 'string'), ('ANO', 'int'), ('MES', 'int'), ('TOTAL', 'bigint')]
>>>
>>> dadosLocacaoDes.dtypes
[('ID', 'int'), ('MODELO', 'string'), ('ANO', 'int'), ('MES', 'int'), ('TOTAL', 'bigint')]
>>>
>>> dadosFaturamento.dtypes
[('ANO', 'int'), ('MES', 'int'), ('TOTAL', 'double')]
>>>
>>> dadosRisco.dtypes
[('ID', 'string'), ('IDCLIENTE', 'string')]
>>>
>>> ■
```

Contratos – Stream e batch (LAMBDA)



Evidências de funcionamento - Hadoop

<https://ssh.cloud.google.com/v2/ssh/projects/engdados-santander/zones/southamerica-central/nodes/namenode>

SSH no navegador

```
[hadoop@namenode ~]$ jps
11265 Jps
11155 NameNode
8675 RunJar
14164 JobHistoryServer
8869 RunJar
11399 SecondaryNameNode
11643 ResourceManager
[hadoop@namenode ~]$ ps -ef | grep hive
hadoop 8675 1 0 May30 ? 00:03:18 /usr/local/java/jdk1.8.0_333/bin/java -Dlog4j.configurationFile=hive-log4j2.properties -Djava.util.logging.config.file=/opt/hive/conf/parquet-logging.properties -Dyarn.log.dir=/opt/hadoop/logs -Dyarn.log.file=hadoop.log -Dyarn.home.dir=/opt/hadoop -Dyarn.root.logger=INFO,console -Djava.library.path=/usr/local/java/jdk1.8.0_333/lib:/opt/hadoop/lib/native -Xmx256m -Dhadoop.log.dir=/opt/hadoop/logs -Dhadoop.log.file=hadoop.log -Dhadoop.home.dir=/opt/hadoop -Dhadoop.id.str=hadoop -Dhadoop.root.logger=INFO,console -Dhadoop.policy.file=hadoop-policy.xml -Dhadoop.security.logger=INFO,NullAppender org.apache.hadoop.util.RunJar /opt/hive/lib/hive-metastore-3.1.2.jar org.apache.hadoop.hive.metastore.HiveMetaStore --hiveconf hive.root.logger=DEBUG,console
hadoop 8869 1 0 May30 ? 00:12:09 /usr/local/java/jdk1.8.0_333/bin/java -Dproc_jar -Dproc_hiveserver2 -Dlog4j.configurationFile=hive-log4j2.properties -Djava.util.logging.config.file=/opt/hive/conf/parquet-logging.properties -Djline.terminal=jline.UnsupportedTerminal -Dyarn.log.dir=/opt/hadoop/logs -Dyarn.log.file=hadoop.log -Dyarn.home.dir=/opt/hadoop -Dyarn.root.logger=INFO,console -Djava.library.path=/usr/local/java/jdk1.8.0_333/lib:/opt/hadoop/lib/native -Xmx256m -Dhadoop.log.dir=/opt/hadoop/logs -Dhadoop.log.file=hadoop.log -Dhadoop.home.dir=/opt/hadoop -Dhadoop.id.str=hadoop -Dhadoop.root.logger=INFO,console -Dhadoop.policy.file=hadoop-policy.xml -Dhadoop.security.logger=INFO,NullAppender org.apache.hadoop.util.RunJar /opt/hive/lib/hive-service-3.1.2.jar org.apache.hive.service.server.HiveServer2 --hiveconf hive.server2.thrift.port=10000 --hiveconf hive.root.logger=DEBUG,console
hadoop 11313 11217 0 03:40 pts/1 00:00:00 grep --color=auto hive
[hadoop@namenode ~]$
```

<https://ssh.cloud.google.com/v2/ssh/projects/engdados-santander/zones/southamerica-central/nodes/datanode>

SSH no navegador

```
[hadoop@datanode ~]$ jps
6496 Jps
25937 DataNode
[hadoop@datanode ~]$
```

Evidências de funcionamento – Hadoop - WebUI

Não seguro | 34.151.243.241:9870/dfshealth.html#tab-overview

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress

Overview 'namenode:9000' (✓active)

Started:	Mon May 30 04:46:19 -0300 2022
Version:	3.3.2, r0bcb014209e219273cb6fd4152df7df713cbac61
Compiled:	Mon Feb 21 15:39:00 -0300 2022 by chao from branch-3.3.2
Cluster ID:	CID-ea2b7f4f-924a-4509-88b7-a8082ecdc620
Block Pool ID:	BP-989863557-10.10.0.217-1653627988109

Summary

Security is off.

Safemode is off.

186 files and directories, 70 blocks (70 replicated blocks, 0 erasure coded blocks)

Heap Memory used 399.96 MB of 675 MB Heap Memory. Max Heap Memory used 675 MB

Non Heap Memory used 82.98 MB of 84.88 MB Committed Non Heap Memory

Não seguro | 34.151.243.241:9870/dfshealth.html#tab-datanode

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

Datanode Information

In service: 1 Down: 0 Decommissioning: 0 Decommissioned: 0 Decommissioned & dead: 0 Entering Maintenance: 0 In Maintenance: 0 In Maintenance & dead: 0

Datanode usage histogram

Disk usage of each DataNode (%)

Não seguro | 34.151.243.241:9870/explorer.html#/

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

Browse Directory

/

Show 25 entries Search:

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	hadoop	supergroup	0 B	May 30 14:33	0	0 B	datalake
drwxrwxrwx	hadoop	supergroup	0 B	May 30 01:13	0	0 B	tmp
drwxr-xr-x	hadoop	supergroup	0 B	May 28 00:39	0	0 B	user

Showing 1 to 3 of 3 entries

Previous 1 Next

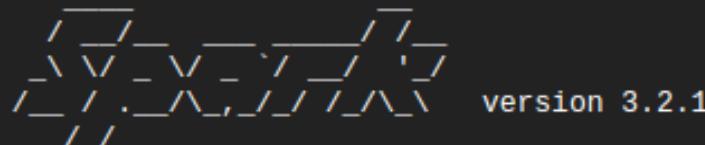
Hadoop, 2022.

Evidências de funcionamento - Spark

<https://ssh.cloud.google.com/v2/ssh/projects/engdados-santander/zones/southamerica-east1-infra/instances/spark-master>

SSH no navegador

```
(base) [spark@master ~]$ jps
2400 Master
12839 SparkSubmit
24329 Jps
(base) [spark@master ~]$ pyspark
Python 3.9.12 (main, Apr 5 2022, 06:56:58)
[GCC 7.5.0] :: Anaconda, Inc. on linux
Type "help", "copyright", "credits" or "license" for more info
Using Spark's default log4j profile: org/apache/spark/log4j-de
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
22/06/01 03:35:11 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using built-in-java classes where applicable
22/06/01 03:35:13 WARN Utils: Service 'SparkUI' could not bind on port 4040. Attempting port 4041.
Welcome to
```



The logo consists of a stylized tree or spark icon made of diagonal lines, followed by the text "version 3.2.1".

```
Using Python version 3.9.12 (main, Apr 5 2022 06:56:58)
Spark context Web UI available at http://master.southamerica-east1-a.c.engdados-santander.internal:4041
Spark context available as 'sc' (master = local[*], app id = local-1654054513162).
SparkSession available as 'spark'.
>>> 
```

<https://ssh.cloud.google.com/v2/ssh/projects/engdados-santander/zones/southamerica-east1-infra/instances/spark-worker>

SSH no navegador

```
[spark@worker ~]$ jps
2275 Worker
11110 Jps
[spark@worker ~]$ 
```

Evidências de funcionamento – Spark Jobs

https://ssh.cloud.google.com/v2/ssh/projects/engdados-santander/zones/southamerica-east1-a/instances/master?authuser=0&hl=pt_BR&pr...

SSH no navegador

```
(base) [spark@master jobs]$ crontab -l
0 10 * * * /opt/spark/bin/spark-submit --deploy-mode client /home/spark/jobs/silver.py
0 11 * * * /opt/spark/bin/spark-submit --deploy-mode client /home/spark/jobs/gold.py
0 9 * * * /opt/spark/bin/spark-submit --deploy-mode client /home/spark/jobs/bronze.py
(base) [spark@master jobs]$ ls
bronze.py enviar_contrato.py get_stream.py gold.py silver.py
(base) [spark@master jobs]$ ps -ef | grep get_stream
spark    12839      1  2 May30 ?        01:20:47 /usr/local/java/jdk1.8.0_333/bin/java -cp /opt/spark/conf/:/opt
/spark/jars/* -Xmx1g org.apache.spark.deploy.SparkSubmit --deploy-mode client --packages org.apache.spark:spark
-sql-kafka-0-10_2.12:3.2.1 /home/spark/jobs/get_stream.py
spark    12876  12839  0 May30 ?        00:00:38 python3 /home/spark/jobs/get_stream.py
spark    24762  24704  0 04:43 pts/0    00:00:00 grep --color=auto get_stream
(base) [spark@master jobs]$
```

Evidências de funcionamento – Spark

The screenshot shows the Apache Spark 3.2.1 Structured Streaming application UI. The top navigation bar includes links for Jobs, Stages, Storage, Environment, Executors, SQL, and Structured Streaming. The Structured Streaming tab is selected. The page title is "pipeline-stream application UI". The main content area is titled "Streaming Query" and displays a table for "Active Streaming Queries (1)". The table has columns for Name, Status, ID, Run ID, Start Time, Duration, Avg Input /sec, Avg Process /sec, and Latest Batch. One row is shown for a query named "<no name>" which is RUNNING, with ID 466401d9-220b-4e41-b4ef-5d48d35b6a4b, Run ID 6c322f08-680d-4c9f-9506-7bb1aa1a925c, started on 2022/05/30 06:58:25, and has been running for 1 day 20 hours 39 minutes. The latest batch is at index 8.

The screenshot shows the Apache Spark 3.2.1 master UI. The top navigation bar includes links for W, Cc, Dc, Cc, In, Bi, La, Pr, Sif, SF, Fl, Fl, Le, In, and spark. The URL in the address bar is "spark://master.southamerica-east1-a.c.engdados-santander.internal:7077". The main content area displays system statistics: URL: spark://master.southamerica-east1-a.c.engdados-santander.internal:7077, Alive Workers: 1, Cores in use: 4 Total, 0 Used, Memory in use: 13.5 GiB Total, 0.0 B Used, Resources in use: Applications: 0 Running, 0 Completed, Drivers: 0 Running, 0 Completed, and Status: ALIVE. Below this, there is a section for "Workers (1)" with a table showing one worker with Worker Id "worker-20220527050347-10.10.0.218-38582", Address "10.10.0.218:38582", State "ALIVE", Cores "4 (0 Used)", and Memory "13.5 GiB (0.0 B Used)".

The screenshot shows the Apache Spark 3.2.1 master UI. It displays sections for "Running Applications (0)" and "Completed Applications (0)". Both sections have tables with columns for Application ID, Name, Cores, Memory per Executor, Resources Per Executor, Submitted Time, User, State, and Duration. The "Running Applications (0)" table is empty. The "Completed Applications (0)" table is also empty.

Evidências de funcionamento - Nifi

The screenshot shows the NiFi web interface with two monitoring cards displayed on the right side of the dashboard.

Mysql

Metric	Value	Time Period
Queued	0 (0 bytes)	
In	0 (0 bytes) → 0	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 → 0 (0 bytes)	5 min

MongoDB

Metric	Value	Time Period
Queued	0 (0 bytes)	
In	0 (0 bytes) → 0	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 → 0 (0 bytes)	5 min

NiFi Flow

Operate

- NiFi Flow Process Group
11b58b85-0181-1000-ce05-e2db1e7b162a
- Actions: Settings, Refresh, Run, Stop, Suspend, Resume, Delete

Metrics

- Nodes: 0
- Bytes: 0 / 0 bytes
- Threads: 0
- Processors: 0
- Connections: 0
- Tasks: 10
- Errors: 0
- Success: 0
- Failure: 0
- Cancel: 0
- Info: 0
- Help: 0

Evidências de funcionamento - Nifi

Não seguro | 34.151.219.59:8080/nifi/?processGroupId=12308566-0181-1000-c6c7-b871cdbe53bc&componentIds=

8080/nifi/ 03:42:41 UTC 03:42:16 UTC

The screenshot shows the Apache NiFi user interface with a process flow. On the left, the 'Operate' sidebar is open, showing a 'MongoDB' process group. The main canvas displays a flow starting with a 'GetMongo' component (QueryMongoDBRecord 1.16.2) which has an 'In' connection to a 'PutHDFS' component (PutHDFS 1.16.2). The 'PutHDFS' component has an 'Out' connection to five 'PutHDFS' components (PutHDFS 1.16.2) for different tables: 'tbl_clientes', 'tbl_veiculos', 'tbl_locacao', 'tbl_despachantes', and another unnamed 'PutHDFS' component. Each of these five PutHDFS components has an 'In' connection to its respective 'QueryDatabaseTableRecord' component ('tbl_clientes', 'tbl_veiculos', 'tbl_locacao', 'tbl_despachantes', and 'tbl_despachantes'). The 'PutHDFS' components also have an 'Out' connection to their respective 'PutHDFS' components. A tooltip 'Name failure, original, success' is shown above the 'PutHDFS' component on the left.

GetMongo
GetMongo 1.16.2
org.apache.nifi - nifi-mongodb-nar

In 0 (0 bytes) 5 min
Read/Write 0 bytes / 0 bytes 5 min
Out 0 (0 bytes) 5 min
Tasks/Time 0 / 0:00:00.000 5 min

Name failure, original, success
Queued 0 (0 bytes)

PutHDFS
PutHDFS 1.16.2
org.apache.nifi - nifi-hadoop-nar

In 0 (0 bytes) 5 min
Read/Write 0 bytes / 0 bytes 5 min
Out 0 (0 bytes) 5 min
Tasks/Time 0 / 0:00:00.000 5 min

tbl_clientes
QueryDatabaseTableRecord 1.16.2
org.apache.nifi - nifi-standard-nar

In 0 (0 bytes) 5 min
Read/Write 0 bytes / 0 bytes 5 min
Out 0 (0 bytes) 5 min
Tasks/Time 0 / 0:00:00.000 5 min

Name success
Queued 0 (0 bytes)

tbl_veiculos
QueryDatabaseTableRecord 1.16.2
org.apache.nifi - nifi-standard-nar

In 0 (0 bytes) 5 min
Read/Write 0 bytes / 0 bytes 5 min
Out 0 (0 bytes) 5 min
Tasks/Time 0 / 0:00:00.000 5 min

Name success
Queued 0 (0 bytes)

tbl_locacao
QueryDatabaseTableRecord 1.16.2
org.apache.nifi - nifi-standard-nar

In 0 (0 bytes) 5 min
Read/Write 0 bytes / 0 bytes 5 min
Out 0 (0 bytes) 5 min
Tasks/Time 0 / 0:00:00.000 5 min

Name success
Queued 0 (0 bytes)

tbl_despachantes
QueryDatabaseTableRecord 1.16.2
org.apache.nifi - nifi-standard-nar

In 0 (0 bytes) 5 min
Read/Write 0 bytes / 0 bytes 5 min
Out 0 (0 bytes) 5 min
Tasks/Time 0 / 0:00:00.000 5 min

Name success
Queued 0 (0 bytes)

PutHDFS
PutHDFS 1.16.2
org.apache.nifi - nifi-hadoop-nar

In 0 (0 bytes) 5 min
Read/Write 0 bytes / 0 bytes 5 min
Out 0 (0 bytes) 5 min
Tasks/Time 0 / 0:00:00.000 5 min

PutHDFS
PutHDFS 1.16.2
org.apache.nifi - nifi-hadoop-nar

In 0 (0 bytes) 5 min
Read/Write 0 bytes / 0 bytes 5 min
Out 0 (0 bytes) 5 min
Tasks/Time 0 / 0:00:00.000 5 min

PutHDFS
PutHDFS 1.16.2
org.apache.nifi - nifi-hadoop-nar

In 0 (0 bytes) 5 min
Read/Write 0 bytes / 0 bytes 5 min
Out 0 (0 bytes) 5 min
Tasks/Time 0 / 0:00:00.000 5 min

PutHDFS
PutHDFS 1.16.2
org.apache.nifi - nifi-hadoop-nar

In 0 (0 bytes) 5 min
Read/Write 0 bytes / 0 bytes 5 min
Out 0 (0 bytes) 5 min
Tasks/Time 0 / 0:00:00.000 5 min

PutHDFS
PutHDFS 1.16.2
org.apache.nifi - nifi-hadoop-nar

In 0 (0 bytes) 5 min
Read/Write 0 bytes / 0 bytes 5 min
Out 0 (0 bytes) 5 min
Tasks/Time 0 / 0:00:00.000 5 min

NiFi Flow > MongoDB

Evidências de funcionamento - Hue

The screenshot displays the Hue web interface, version 3.1.0, running at 34.151.243.241:8000/hue/editor?editor=3. The interface is split into two main sections: the top section shows a query against a MySQL database, and the bottom section shows a query against a Hive metastore.

Top Section (MySQL Query):

- Query:** `1| SELECT * FROM `locadora`.`clientes` LIMIT 100;`
- Logs:** INFO : Concurrency mode is disabled, not creating a lock manager
DEBUG : Shutting down query `SELECT * FROM `locadora`.`clientes` LIMIT 100`
- Table Definition:** Shows the structure of the `locadora.clientes` table.

	Tables	Filter...
1	locadora	(11) + ↻
2	clientes	
3	contratos	
4	contratosrisco	
5	contratosstream	
6	despachantes	
7	faturamento	

Bottom Section (Hive Query):

- Query:** `1| Example: SELECT * FROM ta`
- Logs:** No logs are present.
- Table Definition:** Shows the structure of the `locadora.clientes` table.
- Results:** A table showing the first 6 rows of the `clientes` table.

	clientes.idcliente	clientes.cpf
1	1	d5b84d3fbf1b14b39
2	2	58770104bdbefc2aa
3	3	30d0439454c4e853
4	4	cfc93c5d318963d93
5	5	98e331a488925961
6	6	abfd6c64601ca6b9c

Evidências de funcionamento – Tools - Hive

DBeaver 22.0.5 - clientes

File Edit Navigate Procurar Editor de SQL Banco de dados Window Help

SQL Aplicar Reverter Auto 34.151.243.241 locadora Tabelas clientes

Navegador de banco de Dados Projects

Enter a part of object name here

34.151.243.241 - 34.151.243.241:10000

- default
- locadora
 - Tabelas
 - clientes
 - contratos
 - contratosrisco
 - contratosstream
 - despachantes
 - faturamento
 - locacao
 - locacaoclientes
 - locacaodespachantes

Project - General X

Name Data Source

Bookmarks ER Diagrams

clientes

Properties Dado ER Diagrama

clientes Enter a SQL expression to filter results (use Ctrl+Space)

Grade	123 idcliente	ABC cpf	ABC cnh	ABC validadecnh	ABC nome
1	1	d5b84d3fbf1b14b398501f6759d4b9f	dc9ebec7d697d0a08aa54b82c386127	1632960000000	ANONIM
2	2	58770104bdbefc2aa3defdf9c226a45t	bcf631134a1f7e4707f03d578e5d01fa	1834963200000	ANONIM
3	3	30d0439454c4e853883f691d4436edd	5a27749d79c46483a16fb8867692eb4	1733875200000	ANONIM
4	4	cfc93c5d318963d9382ec18444cc0914	5aa2efc2cc29a358a7a9f0bd8394bf5b	1754092800000	ANONIM
5	5	98e331a488925961c5a317e2fb445fc	9cd3b4967369f40790030b7a701e50d	1597968000000	ANONIM
6	6	abfd6c64601ca6b99aae590dd97d3ea	dad9a77d37d0aa661c352704c351dbt	1591315200000	ANONIM
7	7	6df94746e25837746c1d554d9947566	dc8b907c5343c294493a18cd22dfc7t	1811376000000	ANONIM
8	8	4dcca57b9bf30a6b87b7ea97f2548b1	bb41ae58c95881f63ac9b7c3fca33d58	1778889600000	ANONIM
9	9	cae64880f799138c9698e1edd72a17b	075f536f39ae576ae7ab1e57b01036c9	1827100800000	ANONIM
10	10	8b09db2985c70e17641e2d2c3523f6c	83947078c3df117e0eb955c69fadfdfb	1640390400000	ANONIM
11	11	5b0ffbcf40137193a862b7c5fa8b2507	21bbec32faba87bdfa75f05f2b551b9t	1809302400000	ANONIM
12	12	3ffcf3ccb6bc9dac46f2483fb4ecc	5969a1ee00dbebebb69f97a2a4e5034	1590105600000	ANONIM
13	13	35db2bd775413a594550dc bac90429t	66bbf0ebd10d38ee9318cd0d5c05baa	1616630400000	ANONIM
14	14	de09d43398f3dbd79f3e78c02283be4	1282214b63ffe818e54891d2f10edabt	1772928000000	ANONIM
15	15	9fbe75605da34ead5ce13c1dbd72c7a	8c4ae569624df113a5e8ba973f7db82t	1809561600000	ANONIM
16	16	0b7a8aa13e7ee33d1550d3ae768004t	2c9f4876290a39b4b9dc870179d5716	1847577600000	ANONIM
17	17	44608c671480518c92297938f6898771	b52d3271a3d5d38780d3a2a28bf33bt	1653955200000	ANONIM
18	18	999d58085809b29aa55c92c52cce2cd	2a77b699e541fc1db4cd678455d0ad6	1679616000000	ANONIM
19	19	831198b97c75c20d5de41bb16f32f4et	76688f71d1865a8de66cb503be1665b	1662422400000	ANONIM

Evidências de funcionamento – Tools - Mysql

DBeaver 22.0.5 - CLIENTES

File Edit Navigate Procurar Editor de SQL Banco de dados Window Help

Navegador de banco de Dados Projects

Enter a part of object name here

34.151.243.241 - 34.151.243.241:10000

- default
- locadora

34.95.210.131 - 34.151.219.59:3306

- locadora
- CLIENTES** 16K
- DESPACHANTES 16K
- LOCACAO 16K
- VEICULOS 16K

sys

MySQL - localhost - localhost:3306

Project - General x sys

Name DataSource

Bookmarks ER Diagrams Scripts

Properties Dado ER Diagrama

CLIENTES Enter a SQL expression to filter results (use Ctrl+Space)

Grade	IDCLIENTE	CPF	CNH	VALIDADECNH	NOME	DATACADASTRO
1	1	16528447080	44971318961	2021-09-30	Alvito Espargosa	2019-01-25
2	2	83327244065	44976832700	2028-02-24	Blasco Canto	2019-01-15
3	3	54513828080	27672609233	2024-12-11	Carminda Veríssimo	2019-04-09
4	4	54075418316	75523770439	2025-08-02	Cleusa Lamego	2019-05-26
5	5	41148381120	55079718160	2020-08-21	Clóvis Carrasco	2019-01-28
6	6	85588747194	30471250272	2020-06-05	Cátia Fróis	2019-05-23
7	7	10003401804	45948223160	2027-05-27	Emílio Faro	2019-04-17
8	8	41555470319	15942100562	2026-05-16	Hipólito Granja	2019-05-13
9	9	86405032817	43316295637	2027-11-25	Iara Cardoso	2019-04-04
10	10	15516586747	76742255766	2021-12-25	Isaura Farias	2019-05-04
11	11	44818765309	63408388553	2027-05-03	Iuri Alancastre	2019-02-18
12	12	11257675109	41824436778	2020-05-22	Laura Marcondes	2019-06-30
13	13	88357618405	49548848905	2021-03-25	Micael Mangueira	2019-05-24
14	14	58831402110	84776516129	2026-03-08	Márcio Taveira	2019-02-13
15	15	86405301141	89650561775	2027-05-06	Noêmia Tupinambá	2019-06-28
16	16	70686300220	95360639928	2028-07-19	Paula Padilha	2019-04-09
17	17	70805566600	61460867546	2022-05-31	Rebeca Torcuato	2019-02-09
18	18	84611132765	85761272593	2023-03-24	Rosana Betancour	2019-05-10
19	19	72887644190	40239379806	2022-09-06	Sabino Abrantes	2019-05-27
20	20	73201278572	56448355307	2025-02-02	Severino Leiria	2019-02-15

Rows: 1 25 row(s) fetched - 56ms (11ms fetch), on jun 01, 00:46:07

Evidências de funcionamento – Tools – Mongo e Kafka

Robo 3T - 1.4

File View Options Window Help

Mongo-CaseEngDad... System config locadora Collections (1) contratos Indexes Functions Users

db.getCollection('contra...')

Mongo-CaseEngDados 34.151.219.59:27017 locadora

db.getCollection('contratos').find({})

contratos 0.065 sec.

Key	Value	Type
(1) ObjectId("62968fcfb835f0d57d9ca5d2")	{ 5 fields }	Object
(2) ObjectId("62968fcfb835f0d57d9ca5d5")	{ 5 fields }	Object

ssh.cloud.google.com/v2/ssh/projects/engdados-santander/zones/southamerica-east1-b/instances/tools1?authuser=0&hl=pt_BR&projectNumber=829971108144&useAdminProxy=true&troubles...

SSH no navegador

```
>&1
kafka    903  899  0 May31 ?          00:02:58 java -Xmx1G -Xms1G -server -XX:+UseG1GC -XX:MaxGCPauseMillis=20 -XX:InitiatingHeapOccupancyPercent=35 -XX:+ExplicitGCInvokesConcurrent -XX:MaxInlineLevel=15 -Djava.awt.headless=true -Xloggc:/opt/kafka/bin/../logs/kafkaServer-gc.log -XX:+PrintGCDetails -XX:+PrintGCDateStamps -XX:+PrintGCTimeStamps -XX:+UseGCLogFileRotation -XX:NumberOfGCLogFiles=10 -XX:GCLogFileSize=100M -Dcom.sun.management.jmxremote -Dcom.sun.management.jmxremote.authenticate=false -Dcom.sun.management.jmxremote.ssl=false -Dkafka.logs.dir=/opt/kafka/bin/..../logs -Dlog4j.configuration=file:/opt/kafka/bin/..../conf/log4j.properties -cp /opt/kafka/bin/..../libs/activation-1.1.1.jar:/opt/kafka/bin/..../libs/aopalliance-repackaged-2.6.1.jar:/opt/kafka/bin/..../libs/argparse4j-0.7.0.jar:/opt/kafka/bin/..../libs/audience-annotations-0.5.0.jar:/opt/kafka/bin/..../libs/commons-cli-1.4.jar:/opt/kafka/bin/..../libs/commons-lang3-3.8.1.jar:/opt/kafka/bin/..../libs/connect-api-2.8.1.jar:/opt/kafka/bin/..../libs/connect-basic-auth-extension-2.8.1.jar:/opt/kafka/bin/..../libs/connect-file-2.8.1.jar:/opt/kafka/bin/..../libs/connect-json-2.8.1.jar:/opt/kafka/bin/..../libs/connect-mirror-2.8.1.jar:/opt/kafka/bin/..../libs/connect-mirror-client-2.8.1.jar:/opt/kafka/bin/..../libs/connect-runtime-2.8.1.jar:/opt/kafka/bin/..../libs/connect-transforms-2.8.1.jar:/opt/kafka/bin/..../libs/hk2-api-2.6.1.jar:/opt/kafka/bin/..../libs/hk2-locator-2.6.1.jar:/opt/kafka/bin/..../libs/hk2-utils-2.6.1.jar:/opt/kafka/bin/..../libs/jackson-annotations-2.10.5.jar:/opt/kafka/bin/..../libs/jackson-core-2.10.5.jar:/opt/kafka/bin/..../libs/jackson-databind-2.10.5.1.jar:/opt/kafka/bin/..../libs/jackson-dataformat-csv-2.10.5.jar:/opt/kafka/bin/..../libs/jackson-datatype-jdk8-2.10.5.jar:/opt/kafka/bin/..../libs/jackson-jaxrs-base-2.10.5.jar:/opt/kafka/bin/..../libs/jackson-jaxrs-json-provider-2.10.5.jar:/opt/kafka/bin/..../libs/jackson-module-jaxb-annotations-2.10.5.jar:/opt/kafka/bin/..../libs/jackson-module-paramaner-2.10.5.jar:/opt/kafka/bin/..../libs/jackson-module-scala-2.12-2.10.5.jar:/opt/kafka/bin/..../libs/jakarta.activation-api-1.2.1.jar:/opt/kafka/bin/..../libs/jakarta.annotation-api-1.3.5.jar:/opt/kafka/bin/..../libs/jakarta.inject-2.6.1.jar:/opt/kafka/bin/..../libs/jakarta.validation-api-2.0.2.jar:/opt/kafka/bin/..../libs/jakarta.ws.rs-api-2.1.6.jar:/opt/kafka/bin/..../libs/jakarta.xml.bind-api-2.3.2.jar:/opt/kafka/bin/..../libs/javassist-3.27.0-GA.jar:/opt/kafka/bin/..../libs/javax.servlet-api-3.1.0.jar:/opt/kafka/bin/..../libs/javax.ws.rs-api-2.1.1.jar:/opt/kafka/bin/..../libs/jaxb-api-2.3.0.jar:/opt/kafka/bin/..../libs/jersey-client-2.34.jar:/opt/kafka/bin/..../libs/jersey-common-2.34.jar:/opt/kafka/bin/..../libs/jersey-containter-servlet-2.34.jar:/opt/kafka/bin/..../libs/jersey-server-2.34.jar:/opt/kafka/bin/..../libs/jetty-client-9.4.43.v20210629.jar:/opt/kafka/bin/..../libs/jetty-continuation-9.4.43.v20210629.jar:/opt/kafka/bin/..../libs/jetty-http-9.4.43.v20210629.jar:/opt/kafka/bin/..../libs/jetty-io-9.4.43.v20210629.jar:/opt/kafka/bin/..../libs/jetty-jersey-9.4.43.v20210629.jar:/opt/kafka/bin/..../libs/jetty-server-9.4.43.v20210629.jar:/opt/kafka/bin/..../libs/jetty-util-9.4.43.v20210629.jar:/opt/kafka/bin/..../libs/jetty-util-9.4.43.v20210629.jar:/opt/kafka/bin/..../libs/jetty-servlets-9.4.43.v20210629.jar:/opt/kafka/bin/..../libs/jline-3.12.1.jar:/opt/kafka/bin/..../libs/jopt-simple-5.0.4.jar:/opt/kafka/bin/..../libs/kafka_2.12-2.8.1.jar:/opt/kafka/bin/..../libs/kafka-clients-2.8.1.jar:/opt/kafka-bin/..../libs/kafka-log4j-appender-2.8.1.jar:/opt/kafka/bin/..../libs/kafka-metadata-2.8.1.jar:/opt/kafka/bin/..../libs/kafka-raft-2.8.1.jar:/opt/kafka-bin/..../libs/kafka-shell-2.8.1.jar:/opt/kafka/bin/..../libs/kafka-streams-2.8.1.jar:/opt/kafka/bin/..../libs/kafka-streams-examples-2.8.1.jar:/opt/kafka/bin/..../libs/kafka-streams-scala-2.12-2.8.1.jar:/opt/kafka/bin/..../libs/kafka-streams-test-utils-2.8.1.jar:/opt/kafka/bin/..../libs/kafka-tools-2.8.1.jar:/opt/kafka/bin/..../libs/log4j-2.17.jar:/opt/kafka/bin/..../libs/tz4-java-1.7.1.jar:/opt/kafka/bin/..../libs/maven-artifact-3.8.1.jar:/opt/kafka/bin/..../libs/metrics-core-2.2.0.jar:/opt/kafka/bin/..../libs/netty-buffer-4.1.62.Final.jar:/opt/kafka/bin/..../libs/netty-codec-4.1.62.Final.jar:/opt/kafka/bin/..../libs/netty-common-4.1.62.Final.jar:/opt/kafka/bin/..../libs/netty-handler-4.1.62.Final.jar:/opt/kafka/bin/..../libs/netty-resolver-4.1.62.Final.jar:/opt/kafka/bin/..../libs/netty-transport-4.1.62.Final.jar:/opt/kafka/bin/..../libs/netty-transport-native-epoll-4.1.62.Final.jar:/opt/kafka/bin/..../libs/netty-transport-native-unix-common-4.1.62.Final.jar:/opt/kafka/bin/..../libs/osgi-resource-locator-1.0.3.jar:/opt/kafka/bin/..../libs/paranamer-2.8.jar:/opt/kafka/bin/..../libs/plexus-utils-3.2.1.jar:/opt/kafka/bin/..../libs/reflections-0.9.12.jar:/opt/kafka/bin/..../libs/rocksdbjni-5.18.4.jar:/opt/kafka/bin/..../libs/scala-collection-compat-2.12-2.3.0.jar:/opt/kafka/bin/..../libs/scala-java8-compat-2.12-0.9.1.jar:/opt/kafka/bin/..../libs/scala-library-2.12.13.jar:/opt/kafka/bin/..../libs/scala-logging-2.12-3.9.2.jar:/opt/kafka/bin/..../libs/scala-reflect-2.12.13.jar:/opt/kafka/bin/..../libs/slf4j-api-1.7.30.jar:/opt/kafka/bin/..../libs/slf4j-log4j12-1.7.30.jar:/opt/kafka/bin/..../libs/snappy-java-1.1.8.1.jar:/opt/kafka/bin/..../libs/zookeeper-3.5.9.jar:/opt/kafka/bin/..../libs/zookeeper-jute-3.5.9.jar:/opt/kafka/bin/..../libs/zstd-jni-1.4.9-1.jar kafka.Kafka /opt/kafka/config/server.properties
root    7766  6232  0 04:39 pts/1   00:00:00 su - kafka
kafka   7767  7766  0 04:39 pts/1   00:00:00 -bash
kafka   7817  7767  0 04:39 pts/1   00:00:00 ps -ef
kafka   7818  7767  0 04:39 pts/1   00:00:00 grep --color=auto kafka
[1 kafka@tonisi bin]$
```

Fontes do Case - GitHub

Engenharia de Dados

https://github.com/marciodelima/case-santander_engdados.git

Obrigado