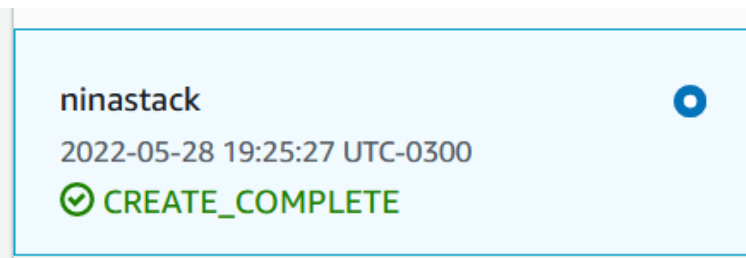


Evidências Sprint AWS + AZURE

☰ Class	Arquitetura em Nuvem PI
🕒 Created	@May 28, 2022 7:28 PM
🔗 Materials	
☑ Reviewed	☑
☰ Type	Resumo

Stack criada



Instâncias

<input type="checkbox"/>	ninastack-sptech	i-0853bdd0e5de77db7	☑ Executando	🔍	t3.small
<input checked="" type="checkbox"/>	fileserver azure	i-08588e1f739139022	☑ Executando	🔍	t2.micro

Buckets

<input type="radio"/>	ninastack-sptech-bucket-bruto	Leste dos EUA (Norte da Virgínia) us-east-1	Bucket e objetos não públicos	28 May 2022 07:25:34 PM -03
<input type="radio"/>	ninastack-sptech-bucket-cliente	Leste dos EUA (Norte da Virgínia) us-east-1	Bucket e objetos não públicos	28 May 2022 07:25:34 PM -03
<input type="radio"/>	ninastack-sptech-bucket-tratados	Leste dos EUA (Norte da Virgínia) us-east-1	Bucket e objetos não públicos	28 May 2022 07:25:34 PM -03

Arquivos (base de dados) gerados e upados no bucket

ninastack-sptech-bucket-bruto [Info](#)

Objetos

Propriedades

Permissões

Métricas


Gerenciamento







Pontos de acesso

Objetos (6)

Os objetos são as entidades fundamentais armazenadas no Amazon S3. Você pode usar o [inventário do Amazon S3](#) para obter um inventário de todos os objetos armazenados no bucket e conceder permissões explicitamente a eles. [Saiba mais](#)

  Copiar URI do S3  Copiar URL  Fazer download  Abrir  Excluir

 Localizar objetos por prefixo

<input type="checkbox"/>	Nome	Tipo
<input type="checkbox"/>	 DICT.xml	xml
<input type="checkbox"/>	 fraudes_pix.json	json
<input type="checkbox"/>	 part-00000-f6a4016b-fc12-4a37-ae9f-24405241c20a-c000.snappy.parquet	parquet
<input type="checkbox"/>	 part-00001-f6a4016b-fc12-4a37-ae9f-24405241c20a-c000.snappy.parquet	parquet
<input type="checkbox"/>	 pessoa.txt	txt
<input type="checkbox"/>	 transacoes_pix.csv	csv

ninastack-sptech-bucket-tratados [Info](#)

Objetos

Propriedades

Permissões

Métricas

Gerenciamento

Objetos (4)

Os objetos são as entidades fundamentais armazenadas no Amazon S3. Você pode usar o [inventário](#) para acessar seus objetos, você precisará conceder permissões explicitamente a eles. [Saiba mais](#)



Copiar URI do S3



Copiar URL



Fazer download



Carregar



Localizar objetos por prefixo



Nome



Tipo



Última modificação



[DICT.txt/](#)

Pasta

-



[fraudes_pix.txt/](#)

Pasta

-



[pessoa.txt/](#)

Pasta

-



[transacoes_pix.txt/](#)

Pasta

-

ninastack-sptech-bucket-cliente [Info](#)

Objetos

Propriedades

Permissões

Métricas

Gerenciamento

P

Objetos (1)

Os objetos são as entidades fundamentais armazenadas no Amazon S3. Você pode usar o [inventário do Am](#) acessarem seus objetos, você precisará conceder permissões explicitamente a eles. [Saiba mais](#) [↗](#)



Copiar URI do S3

Copiar URL

Fazer download



Carregar



Localizar objetos por prefixo



Nome



Tipo



Última modificação



security/

Pasta

-

Servidor + banco sql no Azure

Básico

Assinatura	Azure for Students
Grupo de recursos	resources-nina
Região	East US
Nome do banco de dados	database-gp06-nina
Servidor	(novo) server-gp-nina
Método de autenticação	Autenticação do SQL
Logon do administrador do servidor	adm
Computação + armazenamento	Básico: 2 GB de armazenamento
Redundância do armazenamento de backup	Armazenamento de backup com redundância local

banco de dados azure com a tabela criada + select

Microsoft Azure | database-gp06-nina (server-gp-nina/database-gp06-nina) | Editor de consultas (visualização)

Consulta 1

```
1 select top 10 * from [dbo].[Security]
2
3
```

Resultados

id	VALOR TRANSFERIDO (R\$)	TIPO CHAVE	NOME INSTITUIÇÃO	UF
0	2370407.26	EVP	NU PAGAMENTOS S.A.	RJ
1	6950394.83	CPF	BANCO SANTANDER (BRASL...	SC
5	8053276.03	PHONE	Banco Inter S.A.	AP
6	159.18	EVP	Banco Inter S.A.	MA

Êxito na consulta | 0s

sfileservar no azure

Criar uma conta de armazenamento

✓ Validação aprovada

Básico Avançado Rede Proteção de dados Criptografia Marcas Examinar + criar

Básico

Assinatura	Azure for Students
Grupo de Recursos	resources-nina
Localização	eastus
Nome da conta de armazenamento	fileservnina
Modelo de implantação	Resource manager
Desempenho	Standard
Replicação	RA-GRS (armazenamento com redundância geográfica com acesso de leitura)

replicação dos buckets aws em containers azure

Microsoft Azure

Pesquisar recursos, serviços e documentos (G+/)

Página inicial > fileservernina

fileservernina | Contêineres

Conta de armazenamento

Pesquisar (Ctrl+/)

+ Contêiner Alterar o nível de acesso Restaurar contêineres Atualizar Excluir

Pesquisar contêineres por prefixo

Mostrar contêineres excluídos

Nome	Última modificação	Nível de acesso público	Estado de concessão
<input type="checkbox"/> \$logs	07/06/2022 17:30:39	Privado	Disponível ...
<input type="checkbox"/> containernina	07/06/2022 17:31:57	Privado	Disponível ...
<input type="checkbox"/> containerninacliente	07/06/2022 17:47:11	Privado	Disponível ...
<input type="checkbox"/> containerninatratados	07/06/2022 17:46:59	Privado	Disponível ...

Microsoft Azure

Pesquisar recursos, serviços e documentos (G+/)

Página inicial > fileservernina

containernina

Contêiner

Pesquisar (Ctrl+/)

Carregar Alterar o nível de acesso Atualizar Excluir Alterar a camada Adquirir concessão Interromper concessão Exibir instantâneos

Método de autenticação: Tecla de acesso (Alternar para a Conta de Usuário do Azure AD)
Local: containernina

Pesquisar blobs por prefixo (diferenciar maiúsculas de minúsculas)

Mostrar blobs excluídos

Adicionar o filtro

Nome	Modificado	Camada de acesso	Status do arquivo	Tipo de blob	Tamanho	Estado de concess...
<input type="checkbox"/> DICT.xml	07/06/2022 17:44:44	Principal (Inferidos)		Blob de blocos	42.15 MiB	Disponível ...
<input type="checkbox"/> fraudes_pix.json	07/06/2022 17:44:47	Principal (Inferidos)		Blob de blocos	89.61 MiB	Disponível ...
<input type="checkbox"/> fraudes_pix2.json	07/06/2022 17:44:44	Principal (Inferidos)		Blob de blocos	52.1 MiB	Disponível ...
<input type="checkbox"/> part-00000-f6a4016b-fc12-4a37-ae9f-244052...	07/06/2022 17:44:44	Principal (Inferidos)		Blob de blocos	5.34 MiB	Disponível ...
<input type="checkbox"/> part-00001-f6a4016b-fc12-4a37-ae9f-244052...	07/06/2022 17:44:44	Principal (Inferidos)		Blob de blocos	5.31 MiB	Disponível ...
<input type="checkbox"/> pessoa.txt	07/06/2022 17:44:44	Principal (Inferidos)		Blob de blocos	18.14 MiB	Disponível ...
<input type="checkbox"/> transacoes_pix.csv	07/06/2022 17:44:44	Principal (Inferidos)		Blob de blocos	29.9 MiB	Disponível ...

Microsoft Azure

Pesquisar recursos, serviços e documentos (G+/)

Página inicial > fileservernina

containerninatratados

Contêiner

Pesquisar (Ctrl+/)

Carregar Alterar o nível de acesso Atualizar Excluir Alterar a camada Adquirir concessão Interromper concessão Exibir instantâneos

Método de autenticação: Tecla de acesso (Alternar para a Conta de Usuário do Azure AD)
Local: containerninatratados

Pesquisar blobs por prefixo (diferenciar maiúsculas de minúsculas)

Mostrar blobs excluídos

Adicionar o filtro

Nome	Modificado	Camada de acesso	Status do arquivo	Tipo de blob	Tamanho	Estado de concess...
<input type="checkbox"/> DICT.txt					-	...
<input type="checkbox"/> fraudes_pix.txt					-	...
<input type="checkbox"/> pessoa.txt					-	...
<input type="checkbox"/> transacoes_pix.txt					-	...

Microsoft Azure

Pesquisar recursos, serviços e documentos (G+/)

Página inicial > fileservernina

containerninacliente

Contêiner

Pesquisar (Ctrl+/)

Carregar Alterar o nível de acesso Atualizar Excluir Alterar a camada Adquirir concessão Interromper concessão Exibir instantâneos

Método de autenticação: Tecla de acesso (Alternar para a Conta de Usuário do Azure AD)
Local: containerninacliente

Pesquisar blobs por prefixo (diferenciar maiúsculas de minúsculas)

Mostrar blobs excluídos

Adicionar o filtro

Nome	Modificado	Camada de acesso	Status do arquivo	Tipo de blob	Tamanho	Estado de concess...
<input type="checkbox"/> security					-	...

evidencias pyspark

```
jupyter CONNECTION BUCKET S3 E SQLSERVER AZURE NINA Last checkpoint 6 minutos atrás (unsaved changes) Python 3 (ipykernel)

In [1]: #iniciando spark
from pyspark import SparkConf
from pyspark.sql import SparkSession
from pyspark.sql.functions import col, concat
from pyspark.sql.types import IntegerType
from pyspark.sql import SparkSession
from pyspark.sql.functions import monotonically_increasing_id

conf = SparkConf()
conf.set('spark.jars.packages', 'org.apache.hadoop:hadoop-aws:3.2.2')
conf.set('spark.hadoop.fs.s3a.aws.credentials.provider', 'com.amazonaws.auth.InstanceProfileCredentialsProvider')
spark = SparkSession.builder.config(conf=conf).getOrCreate()

WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.spark.unsafe.Platform (file:/usr/local/lib/python3.7/site-packages/pyspark/jar/s/spark-unsafe-2.12-3.2.1.jar) to constructor java.nio.DirectByteBuffer(long,int)
WARNING: Please consider reporting this to the maintainers of org.apache.spark.unsafe.Platform
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: all illegal access operations will be denied in a future release

;; loading settings :: url = jar:file:/usr/local/lib/python3.7/site-packages/pyspark/jars/ivy-2.5.0.jar!/org/apache/ivy/core/settings/ivysettings.xml

Ivy Default Cache set to: /root/.ivy2/cache
The jars for the packages stored in: /root/.ivy2/jars
org.apache.hadoop:hadoop-aws added as a dependency
com.microsoft.azure:spark-mssql-connector_2.12 added as a dependency
;; resolving dependencies :: org.apache.spark:spark-sql-kernel-parent-4045950-ae64-429d-9052-60a32b0eb41:0
confs: [default]
found org.apache.hadoop:hadoop-aws:3.2.2 in central
found com.amazonaws:aws-java-sdk-bundle:1.11.563 in central
found com.microsoft.azure:spark-mssql-connector_2.12:2.0 in central
found com.microsoft.sqlserver:mssql-jdbc:8.4.1-jre8 in central

In [2]: dfCsv = spark.read.option("delimiter", ";").option("header", "true").csv('s3a://ministack-sptech-bucket-bruto/transacoes_pix.csv')
22/06/08 05:08:05 WARN MetricsConfig: Cannot locate configuration: tried hadoop-metrics2-s3a-file-system.properties,hadoop-metr-ic2.properties

In [ ]: dfCsv.show(vertical=True)

In [3]: _dfCsv = dfCsv.select(
    col('valor').alias('VALOR TRANSFERIDO (R$)'),
    col('tipochave').alias('TIPO CHAVE'),
    col('nomeinstituido').alias('NOME INSTITUICAO'),
    col('siglaestado').alias('UF')
)
# _dfCsv.show()

In [4]: dfCsvSingleCol = _dfCsv.select(concat(*_dfCsv.columns).alias('data'))
# dfCsvSingleCol.show()

In [ ]: _dfCsvSingleCol.write.mode('append').text('s3a://ministack-sptech-bucket-bruto/transacoes_pix.txt')

In [5]: dfTxt = spark.read.option("delimiter", ";").option("header", "true").csv('s3a://ministack-sptech-bucket-bruto/pessoa.txt')

In [ ]: dfTxt.show()

In [6]: _dfTxt = dfTxt.select(
    col('idade').cast(IntegerType()).alias('IDADE'),
)
# _dfTxt.show()

++++++
| IDADE |
++++++
| 24 |
| 44 |
| 20 |
| 45 |
| 30 |
| 36 |
| 34 |
| 37 |
| 45 |
| 33 |
| 48 |
| 18 |
| 23 |
| 38 |
| 21 |
| 18 |
| 28 |
| 22 |
| 28 |
| 36 |
++++++
only showing top 20 rows

In [7]: dfTxtSingleCol = _dfTxt.select(concat(*_dfTxt.columns).alias('data'))
# dfTxtSingleCol.show()

In [ ]: _dfTxtSingleCol.write.mode('append').text('s3a://ministack-sptech-bucket-bruto/pessoa.txt')

In [8]: dfJson = spark.read.json('s3a://ministack-sptech-bucket-bruto/fracoes_pix.json')

In [ ]: dfJson.show(vertical = True)

In [9]: _dfJson = dfJson.select(
    col('status_transacao').alias('STATUS DA TRANSAÇÃO')
)
# _dfJson.show()

In [10]: dfJsonSingleCol = _dfJson.select(concat(*_dfJson.columns).alias('data'))
# dfJsonSingleCol.show()

In [ ]: _dfJsonSingleCol.write.mode('append').text('s3a://ministack-sptech-bucket-bruto/fracoes_pix.txt')

In [11]: import pandas as pd
pdfxml = pd.read_xml('s3a://ministack-sptech-bucket-bruto/DICT.xml')

In [12]: dfXml = spark.createDataFrame(pdfxml)

In [13]: _dfXml = dfXml.select(
    col('tipocontr').alias('TIPO CONTRA'),
    col('ruvidindicadaoacao').alias('CHAVE REIVINDICADA POR DOACAO'),
)
# _dfXml.show()

In [14]: dfXmlSingleCol = _dfXml.select(concat(*_dfXml.columns).alias('data'))

In [ ]: _dfXmlSingleCol.write.mode('append').text('s3a://ministack-sptech-bucket-bruto/DICT.txt')

In [15]: _dfCsv = _dfCsv.withColumn("id", monotonically_increasing_id())

In [16]: _dfTxt = _dfTxt.withColumn("id", monotonically_increasing_id())

In [17]: _dfJson = _dfJson.withColumn("id", monotonically_increasing_id())

In [18]: _dfXml = _dfXml.withColumn("id", monotonically_increasing_id())

In [19]: df_seguranca = _dfCsv.join(_dfTxt, on="id", how="outer").join(_dfJson, on="id", how="outer").join(_dfXml, on="id", how="outer")

In [ ]: # df_seguranca.write.mode('append').text('s3a://ministack-sptech-bucket-bruto/seguranca.txt')

In [ ]: # df_seguranca.write.option("header", "true").csv('s3a://ministack-sptech-bucket-cliente/security', emptyValue='')

In [21]: df_seguranca.write \
    .format('com.microsoft.sqlserver.jdbc.spark') \
    .mode('overwrite') \
    .option('driver', 'com.microsoft.sqlserver.jdbc.SQLServerDriver') \
    .option('url', 'jdbc:sqlserver://server-gp-nina.database.windows.net;database=database-gp8-nina') \
    .option('database', 'dbo.Security') \
    .option('user', 'adm') \
    .option('password', '@nub10n') \
    .save()

22/06/08 05:13:08 WARN TaskSetManager: Stage 7 contains a task of very large size (5313 KiB). The maximum recommended task size is 1000 KiB.

In [22]: import socket
print(socket.gethostname(socket.gethostname()))

10.100.0.245
```