



Inovação com dados em nuvem

TRILHA

#TheDevConf
Oracle

Crie seu pipeline de dados sem pagar por
isso!

Erika Nagamine

23.03.21 12h00



Inovação com dados em nuvem

TRILHA

#TheDevConf
Oracle



Este trabalho está licenciado sob uma Licença Creative Commons Atribuição-Compartilhual 4.0 Internacional.
Para ver uma cópia desta licença, visite <http://creativecommons.org/licenses/by-sa/4.0/>.



Inovação com dados em nuvem

**CRIE SEU PIPELINE DE DADOS
NA CLOUD SEM PAGAR
NADA POR ISSO!**



Índice

| | | |
|------|---|----|
| 1. | Considerações iniciais e pré requisitos | 4 |
| | Recursos usados: | 4 |
| | Tópicos não cobertos: | 4 |
| 2. | Provisionar os recursos | 5 |
| 2.1. | Object Storage | 5 |
| 2.2. | Criando funções | 7 |
| 2.3. | Criando uma evento | 13 |
| 2.4. | Banco de dados Autonomous | 15 |
| 3. | Configurando seu ambiente | 17 |
| 4. | Testando seu pipeline de dados | 22 |
| 5. | Referências | 30 |

1. Considerações iniciais e pré requisitos

A técnica demonstrada aqui no presente documento é apenas uma das que pode ser utilizada para criar um pipeline de dados. É possível utilizar outras técnicas dependendo do seu caso de uso. Para deixar o passo a passo de uma maneira única, padronizamos as telas em inglês.

Recursos usados:

OCI (all free tier)

- Armazenamento Oracle Object storage
- Functions (30 days free)
- Events (30 days free)
- Pacotes Oracle Instant Client + Tools e sqlplus
- Banco de dados autonomous (adw)
- Compute VM 2.1 Shape

Local

- Navegador

Tópicos não cobertos:

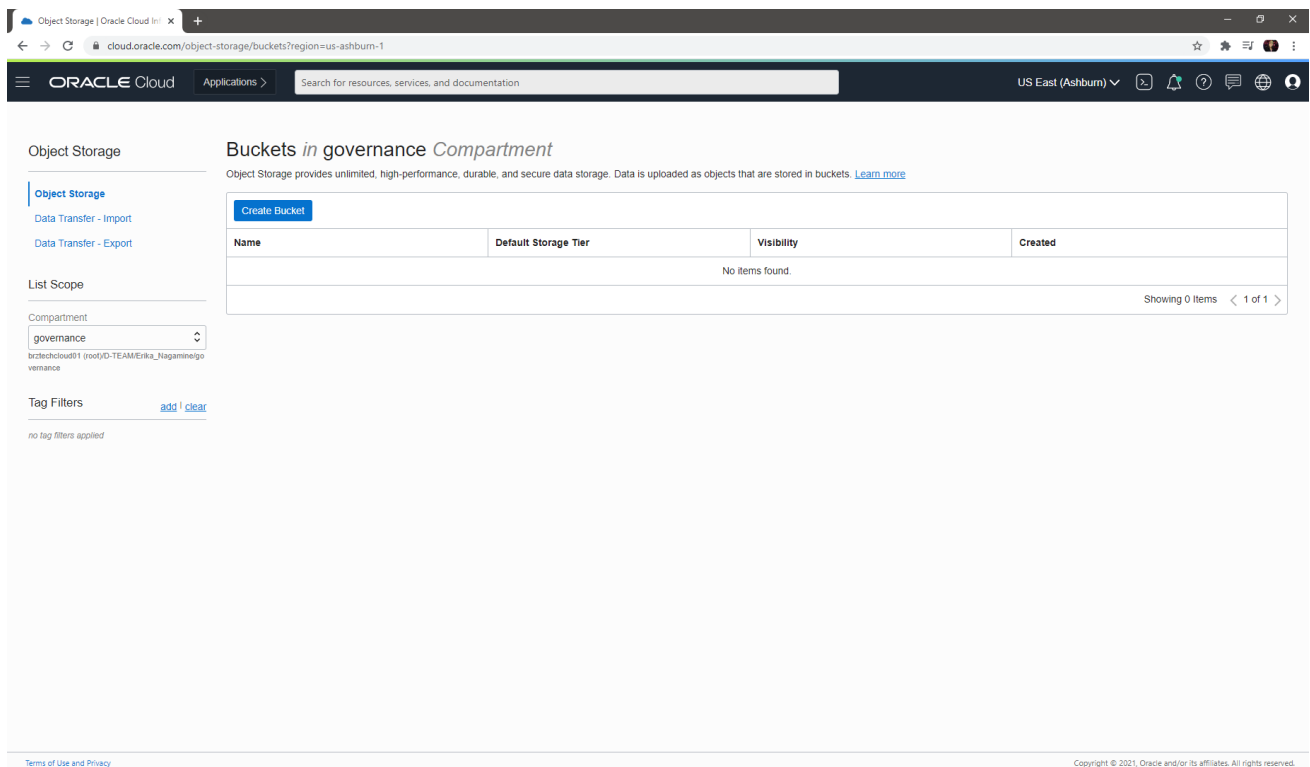
Configuração da máquina adicional para criação de funções

- É possível criar funções através de um computador Always Free, mas não cobriremos essa possibilidade no documento
- Não focaremos na visualização do dados que dentro do serviço de OAC (Oracle analytics Cloud), Oracle Machine Learning Notebooks presente no autonomous ou um relatório via APEX
- A infraestrutura completa também pode ser provisionada via infraestrutura como um código
- Temos por pressuposto que você possui acesso a cloud não cobriremos o tópico de fundação e boas práticas na cloud, como criação de compartments, etc.

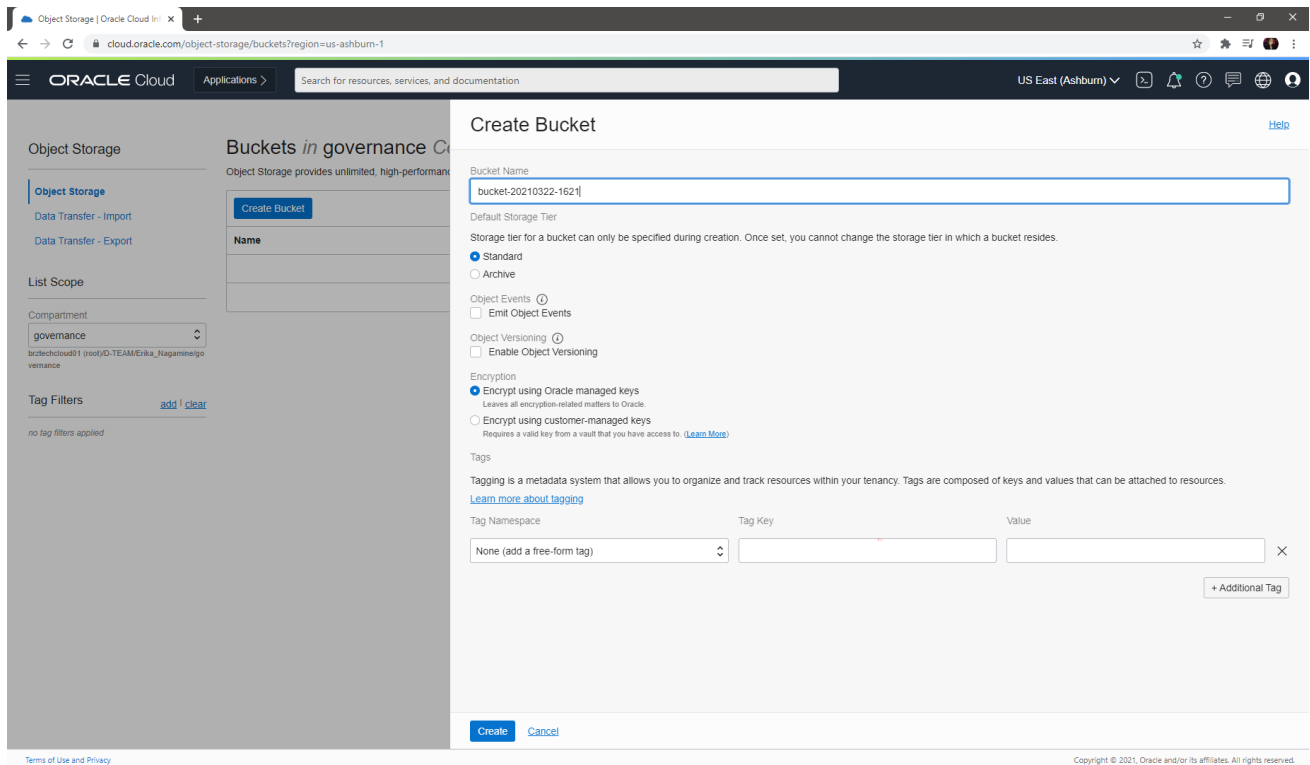
2. Provisionar os recursos

2.1.Object Storage

O object storage é um disco elástico para armazenamento dos mais diversos tipos de arquivos e cargas de trabalho: pode armazenar desde um dado não estruturado até um backup. Para criar um object storage, vá em menu -> Object storage -> object storage:



Iremos criar dois buckets onde um servirá para os dados processados e o outro para dados inseridos (onde vamos receber as informações). Na tela, clique na opção de “create bucket”.



Crie dois buckets com as seguintes informações:

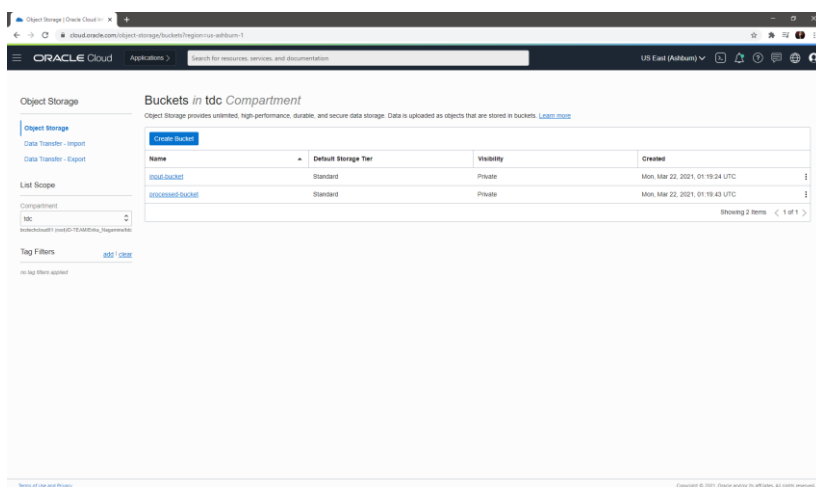
(bucket de entrada de dados)

- Bucket-name: input-bucket
- Object Events: Marque a opção para emitir eventos

(bucket de saída de dados)

- Bucket-name: processed-bucket
- Object Events: Marque a opção para emitir eventos

Ao final dos dois processos, você verá uma tela semelhante:

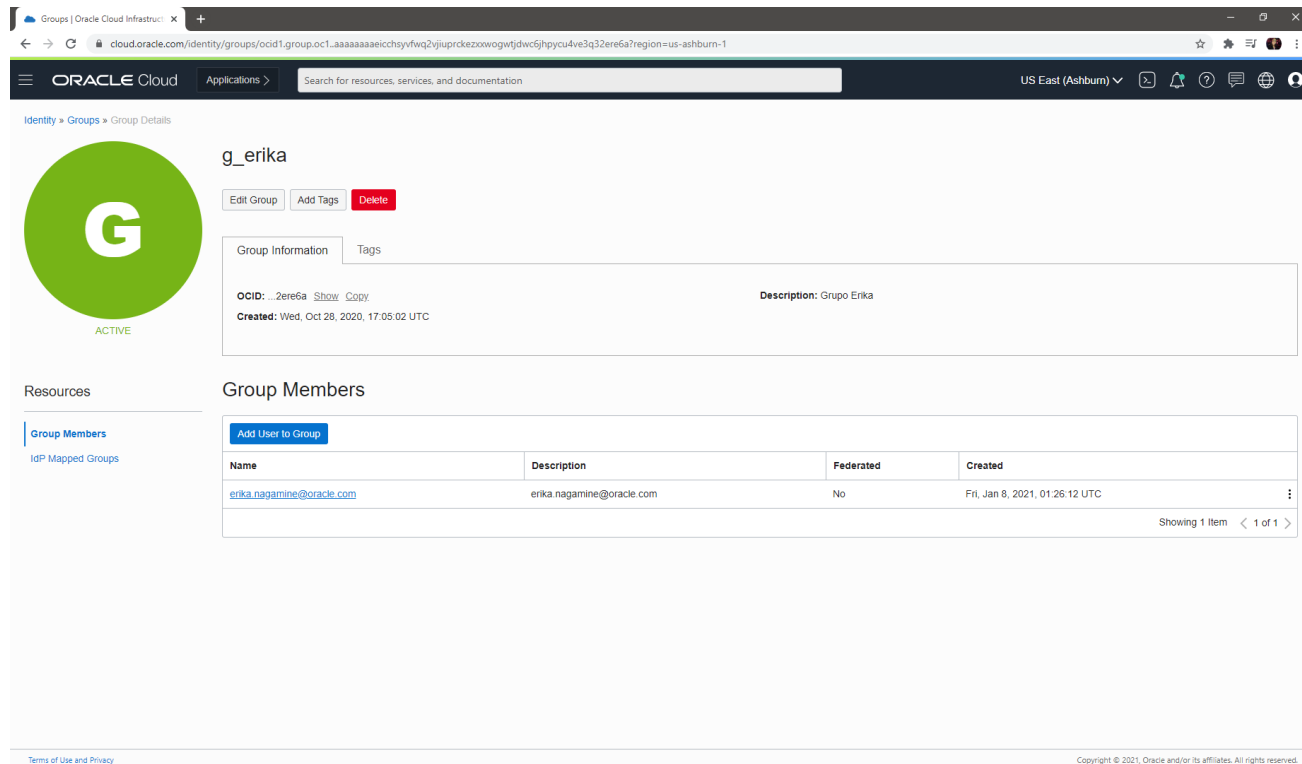


2.2. Criando funções

Alguns pré requisitos deverão ser realizados para esta fase:

- Criação de grupos

Em menu -> identity -> groups crie um grupo para atribuir as políticas necessárias a criação de funções. Um exemplo desta criação pode ser este:

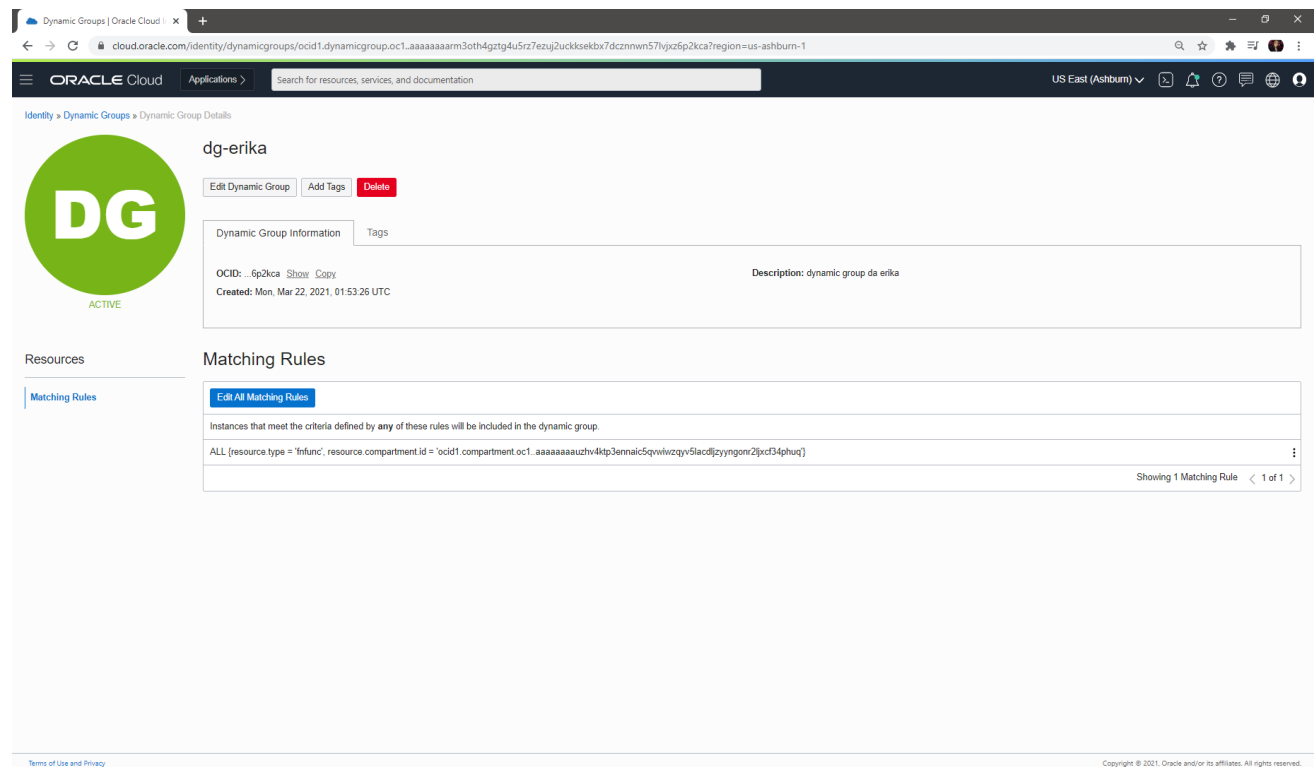


Atribua todos os usuários que precisam ter acesso a este grupo.

- Crie um grupo dinâmico

Para atribuição de permissões sobre seu compartment específico, utilize o recurso de grupo dinâmico. Para cria-lo vá em menu -> identity -> Dynamic Group

```
ALL {resource.type = 'fnfunc', resource.compartment.id =  
'id_do_compartment' }
```

- Criação de políticas
Em menu -> identity – Policies crie as políticas necessárias para esta fase. As políticas necessárias são descritas abaixo:

```
Allow group g_erika to use cloud-shell in tenancy  
Allow group g_erika to manage repos in tenancy  
Allow group g_erika to read objectstorage-namespaces in tenancy  
Allow group g_erika to read metrics in tenancy  
Allow group g_erika to manage functions-family in tenancy  
Allow group g_erika to use virtual-network-family in tenancy  
Allow dynamic-group dg-erika to manage objects in compartment <seu  
compartment> where target.bucket.name=input-bucket  
Allow dynamic-group dg-erika to manage objects in compartment <seu  
compartment> where target.bucket.name=processed-bucket
```

The screenshot shows the Oracle Cloud Identity Policies page for a policy named 'pol_erika'. The policy is active and has a green circular icon with a white 'P'. The page includes a sidebar with 'Resources' and 'Statements' sections. The 'Statements' section lists 11 statements, including permissions for cloud-shell, repos, objectstorage-namespaces, metrics, functions-family, virtual-network-family, and dynamic-group dg-erika to manage objects in tenancy.

Policy Information

- OCID: ...zncgs4a [Show](#) [Copy](#)
- Compartment: brttechcloud01 (root)
- Description: erika
- Created: Mon, Feb 15, 2021, 01:31:56 UTC

Statements

- Allow group g_erika to use cloud-shell in tenancy
- Allow group g_erika to manage repos in tenancy
- Allow group g_erika to read objectstorage-namespaces in tenancy
- Allow group g_erika to read metrics in tenancy
- Allow group g_erika to manage functions-family in tenancy
- Allow group g_erika to use virtual-network-family in tenancy
- Allow dynamic-group dg-erika to manage objects in tenancy where target.bucket.name = 'input-bucket'
- Allow dynamic-group dg-erika to manage objects in tenancy where target.bucket.name = 'processed-bucket'

Showing 11 items

- Criação de VCN

Neste documento criaremos uma rede padrão via “wizard” mas como padrão e boa pratica, pense sempre na boa fundação do seus recursos em cloud. Para criar a rede padrão, vá em menu -> networking -> start VCN Wizard e siga o passo a passo:

The screenshot shows the Oracle Cloud Networking page. The left sidebar lists various networking services: Overview, Virtual Cloud Networks, Dynamic Routing Gateway, Customer-Premises Equipment, VPN Connections, Load Balancers, FastConnect, IP Management, and DNS Management. The main content area is titled 'Networking' and includes a section for 'Networking, Connectivity, and Edge Services' with a 'GET STARTED' button and a 'Create Virtual Cloud Network' button. Below this, there are three 'Create Networking Solutions' cards: 'Create a VCN with Internet Connectivity', 'Create a VCN with Internet Connectivity and VPN Connect', and 'Manage your DNS services'. The right sidebar shows 'Service Health' (All networking systems operational) and 'Documentation' links.

Networking, Connectivity, and Edge Services

Your virtual cloud networks (VCNs) are private, customizable networks that you control. FastConnect, VPN Connect, and Edge Services let you securely connect your VCNs with other private or public networks.

GET STARTED

[Create Virtual Cloud Network](#)

Create Networking Solutions

- Create a VCN with Internet Connectivity**
Creates a VCN with a public subnet that can be reached from the internet. Also creates a private subnet that can connect to the internet through a NAT gateway, and also privately connect to the Oracle Services Network.
[Start VCN Wizard](#) [Show Details and Diagram](#)
- Create a VCN with Internet Connectivity and VPN Connect**
Creates a VCN and sets up VPN Connect to your on-premises network. The VCN also has a public subnet that can be reached from the internet.
[Start VPN Wizard](#) [Show Details and Diagram](#)
- Manage your DNS services**
Manage your DNS Zones, DNS Views, TSIG Keys, Traffic Management Steering Policies, and HTTP Redirects.
[Manage DNS](#)

Service Health

All networking systems operational
[View Health Dashboard](#)

Documentation

- [Networking Overview](#)
- [Virtual Cloud Networks](#)
- [VPN Connect](#)
- [FastConnect](#)
- [Troubleshooting](#)
- [DNS Management](#)

Create a VCN with Internet Connectivity

Configuration

Important: Before starting:

- **Limits:** Ensure your tenancy has not reached its VCN limit. See [Service limits](#)
- **Access:** Ensure you have permission to work in the compartment you select.

Basic Information

VCN NAME ¹

COMPARTMENT ¹

brztechcloud01 (root)

Configure VCN and Subnets

VCN CIDR BLOCK ¹

10.0.0.0/16

If you plan to peer this VCN with another VCN, the VCNs must not have overlapping CIDRs. [Learn more.](#)

PUBLIC SUBNET CIDR BLOCK ¹

10.0.0.0/24

The subnet CIDR blocks must not overlap.

PRIVATE SUBNET CIDR BLOCK ¹

10.0.1.0/24

The subnet CIDR blocks must not overlap.

DNS RESOLUTION

☒ USE DNS HOSTNAMES IN THIS VCN

Required for instance hostname assignment if you plan to use VCN DNS or a third-party DNS. This choice cannot be changed after the VCN is created. [Learn more.](#)

[Show Tagging Options](#)

VCN with Internet Connectivity

Includes:

- VCN
- Public subnet
- Private subnet
- Internet gateway (IG)
- NAT gateway (NAT)
- Service gateway (SG)

Ao final do processo você terá duas subnets:

VCN

AVAILABLE

vcn_enagamine

Move Resource Add Tags Terminate

VCN Information Tags

Compartment: network

Created: Fri, Nov 13, 2020, 04:53:24 UTC

CIDR Block:

OCID: ...laspea [Show](#) [Copy](#)

DNS Resolver: [vcn_enagamine](#)

Default Route Table: [Default Route Table for vcn_enagamine](#)

DNS Domain Name: vcnenagamine.oraclecloud.com

Resources

Subnets (2)

[Subnets \(2\)](#)

[CIDR Blocks \(1\)](#)

[Route Tables \(2\)](#)

[Internet Gateways \(1\)](#)

[Dynamic Routing Gateways \(0\)](#)

[Network Security Groups \(1\)](#)

[Security Lists \(2\)](#)

[DHCP Options \(1\)](#)

[Local Peering Gateways \(0\)](#)

Subnets in network Compartment

Create Subnet

| Name | State | CIDR Block | Subnet Access | Created |
|--|-----------|------------|--------------------|---------------------------------|
| Private Subnet-vcn_enagamine | Available | | Private (Regional) | Fri, Nov 13, 2020, 04:53:29 UTC |
| Public Subnet-vcn_enagamine | Available | | Public (Regional) | Fri, Nov 13, 2020, 04:53:27 UTC |

Showing 2 items < 1 of 1 >

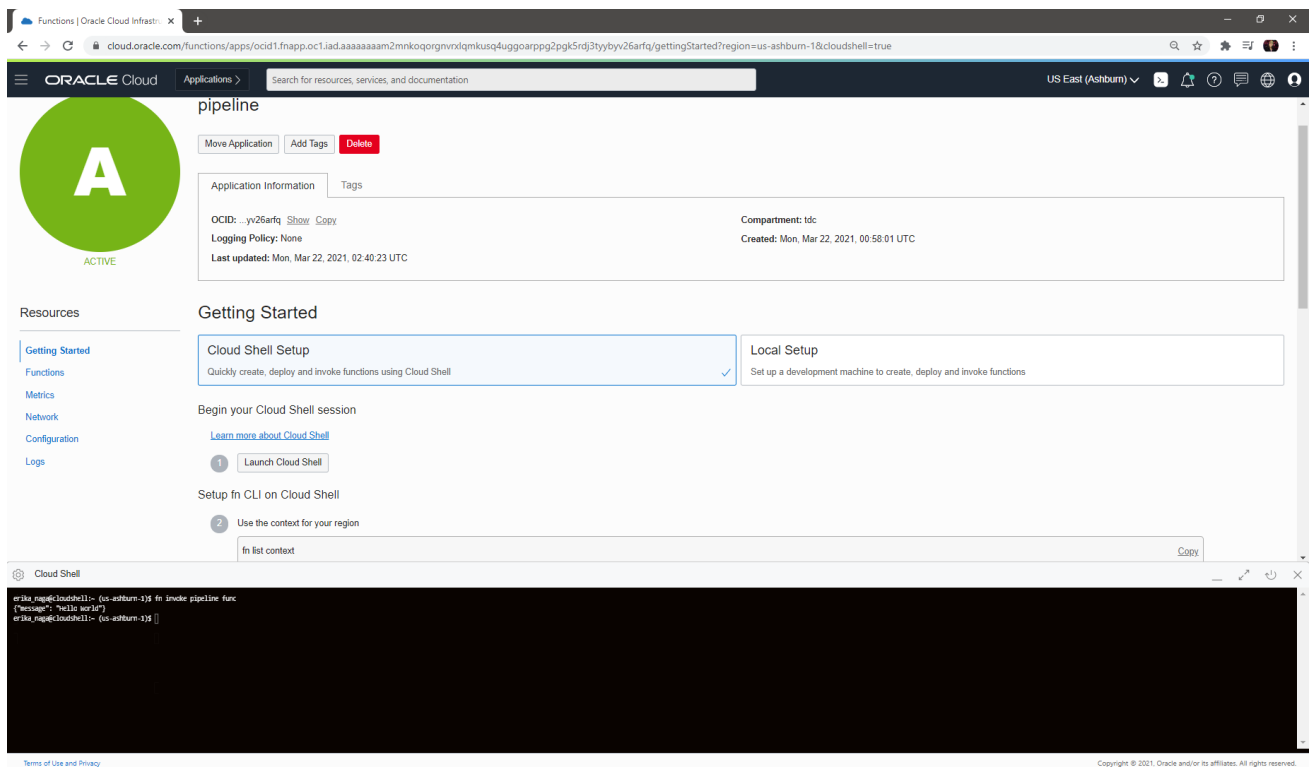
Agora estamos pronto para o processo de criação de funções. Para isso, vá em menu -> developer services -> functions e clique em create application:

The screenshot shows the 'New Application' form in the Oracle Cloud console. The form is titled 'New Application' and is located in the 'Applications in network Compartments' section. The form includes fields for 'Name', 'VCN in network' (with a link to 'Change Compartment'), 'Subnets in network' (with a link to 'Change Compartment'), and 'Tagging' information (Tag Namespace, Tag Key, and Value). The 'Tagging' section has a note: 'Tagging is a metadata system that allows you to organize and track resources within your tenancy. Tags are composed of keys and values that can be attached to resources. Learn more about tagging'. The form also has a 'Create' button and a 'Cancel' button.

Você precisará informar sua VCN criada. Sugerimos (para efeito de workshop, não de dia a dia, colocar na subnet publica).

The screenshot shows the 'pipeline' page in the Oracle Cloud console. The page is titled 'pipeline' and is located in the 'Functions > Applications > pipeline' section. The page includes a 'Move Application' button, an 'Add Tags' button, and a 'Delete' button. The 'Application Information' tab is selected, showing the OCID: yv26arf, Logging Policy: None, and Last updated: Mon, Mar 22, 2021, 02:40:23 UTC. The 'Getting Started' section is expanded, showing 'Cloud Shell Setup' as the selected option. The 'Cloud Shell Setup' section includes a 'Launch Cloud Shell' button and a 'Setup fn CLI on Cloud Shell' section. The 'Setup fn CLI on Cloud Shell' section includes four steps: 1. Use the context for your region, 2. Update the context with the function's compartment ID, 3. Update the context with the location of the Registry you want to use, and 4. Update the context with the location of the Registry you want to use. Each step includes a code snippet and a 'Copy' button.

Após criar sua aplicação para a função, agora vamos configurá-la. Para isso faremos o setup via cloud shell. Siga o passo a passo da página e não se esqueça de gerar e anotar o token:



Agora que a configuração básica da sua função está realizada, no cloud shell baixe os scripts abaixo e coloque os no diretório func:

```
wget https://raw.githubusercontent.com/oracle/oracle-functions-samples/master/samples/oci-load-file-into-adw-python/func.py

wget https://raw.githubusercontent.com/oracle/oracle-functions-samples/master/samples/oci-load-file-into-adw-python/requirements.txt

wget https://raw.githubusercontent.com/oracle/oracle-functions-samples/master/samples/oci-load-file-into-adw-python/func.yaml
```

Faça o deploy da sua função com novas parametrizações:

```
fn -v deploy --app pipeline func
```

Retornaremos a este passo posteriormente para configurar os parâmetros que sua função irá receber.

2.3. Criando uma evento

Navegue até menu -> application integration -> event service

The screenshot shows the Oracle Cloud Events Service console. The main heading is 'Rules in data Compartment'. Below it, a message states: 'Rules trigger actions for selected event conditions. [Learn more.](#)' There is a 'Create Rule' button. A table with columns 'Name', 'Description', 'State', and 'Created Date' is shown, with a message: 'There are no Rules in the data compartment that match the filter criteria.' The bottom of the table shows 'Showing 0 items' and 'Page 1'. The left sidebar has 'Events' as the main section, with 'Rules' and 'Metrics' as sub-sections. Below that is 'List Scope' with a 'Compartment' dropdown set to 'data' and a 'Filters' section with a 'State' dropdown set to 'Any state'. The top navigation bar shows 'ORACLE Cloud' and a search bar. The bottom of the page shows 'Cloud Shell' and 'Terms of Use and Privacy'.

Clique em criar uma nova regra:

The screenshot shows the 'Create Rule' form in the Oracle Cloud Events Service console. The form has several sections: 'Display Name' with a text input field, 'Description' with a text input field, 'Rule Conditions' with a section titled 'Limit the events that trigger actions by defining conditions based on event types, attributes, and filter tags. [Learn more](#)'. This section contains three dropdowns: 'Event Type', 'Service Name', and 'Event Type'. There is a '+ Another Condition' button. To the right of these dropdowns is a 'Rule Logic' section with a text input field and a 'Validate Rule' button. Below the 'Rule Conditions' section is the 'Actions' section, titled 'Actions trigger for the specified event conditions. [Learn more](#)'. It contains a dropdown for 'Action Type' and a '+ Another Action' button. At the bottom of the form are 'Create Rule' and 'Cancel' buttons. The top navigation bar shows 'ORACLE Cloud' and a search bar. The bottom of the page shows 'Cloud Shell' and 'Terms of Use and Privacy'.

Utilize as parametrizações abaixo:

Edit Rule

Display Name: meu_pipeline

Description: meu pipeline em tempo real

Rule Conditions

Limit the events that trigger actions by defining conditions based on event types, attributes, and filter tags. [Learn more](#)

Condition 1: Event Type: Object Storage, Attribute: tdc, Attribute Values: tdc

Condition 2: Attribute: bucketName, Attribute Values: input.bucket

Actions

Actions trigger for the specified event conditions. [Learn more](#)

Action 1: Functions, Function: func

Action 2: Functions, Function: oci-load-file-into-adv-python

Rule Logic

```
MATCH event WHERE (
  eventType EQUALS ANY OF (
    com.oraclecloud.objectstorage.createobject
  )
) AND (
  compartmentName MATCHES ANY OF (
    tdc
  )
) bucketName MATCHES ANY OF (
  input.bucket
)
```

[View example events \(JSON\)](#)

[Save changes](#) [Cancel](#)

Salve as configurações:

meu_pipeline

meu pipeline em tempo real

[Edit Rule](#) [Validate Rule](#) [Disable](#) [Move Resource](#) [More Actions](#)

Details **Tags**

Rule Information

OCID: axi5cfyza [Show](#) [Copy](#)

Compartment: brztechcloud01 (root)ID-TEAMERika_Nagamine:tdc

Created: Mon, Mar 22, 2021, 01:31:41 UTC

Resources

Metrics

Start time: Mar 22, 2021 7:08:40 PM, End time: Mar 22, 2021 8:08:40 PM, Quick Selects: Last hour

Events Delivered

Interval: 1 minute, Statistic: Sum

Delivery Failure

Interval: 1 minute, Statistic: Sum

Nesse momento seu data lake está pronto para receber os dados. Faremos a configuração do nosso data warehouse

2.4. Banco de dados Autonomous

Nosso Data warehouse será construído em cima do autonomous database. Navegue no menu direito até autonomous database e clique em criar autonomous data warehouse.

Em autonomous data warehouse clique em “create autonomous database:

Create Autonomous Database

Provide basic information for the Autonomous Database

Compartment: governance

Display name: DB 202103221723

Database name: DB202103221723

Choose a workload type

- Data Warehouse** (Selected): Built for decision support and data warehouse workloads. Fast queries over large volumes of data.
- Transaction Processing: Built for transactional workloads. High concurrency for short-running queries and transactions.
- JSON: Built for JSON-centric application development. Developer-friendly document APIs and native JSON storage.
- APEX: Built for Oracle APEX application development. Creation and deployment of low-code applications, with database included.

Choose a deployment type

- Shared Infrastructure** (Selected): Run Autonomous Database on shared Exadata Infrastructure.
- Dedicated Infrastructure: Run Autonomous Database on dedicated Exadata Infrastructure.

Configure the database

Choose database version: 19c

OCPU count: 1

Storage (TB): 1

☒ Auto scaling

Create Autonomous Database Cancel

Create Autonomous Database

☒ Auto scaling

Create administrator credentials

Username: ADMIN

Password:

Confirm password:

Choose network access

Access Type

- Allow secure access from everywhere** (Selected): You can restrict access to specific IP addresses and VCNs.
- Virtual cloud network: Private access only, using a VCN.

☐ Configure access control rules

Choose a license type

- Bring Your Own License (BYOL)** (Selected): Bring my organization's Oracle Database software licenses to the Database service.
- License Included: Subscribe to new Oracle Database software licenses and the Database service.

[Show Advanced Options](#)

Create Autonomous Database Cancel

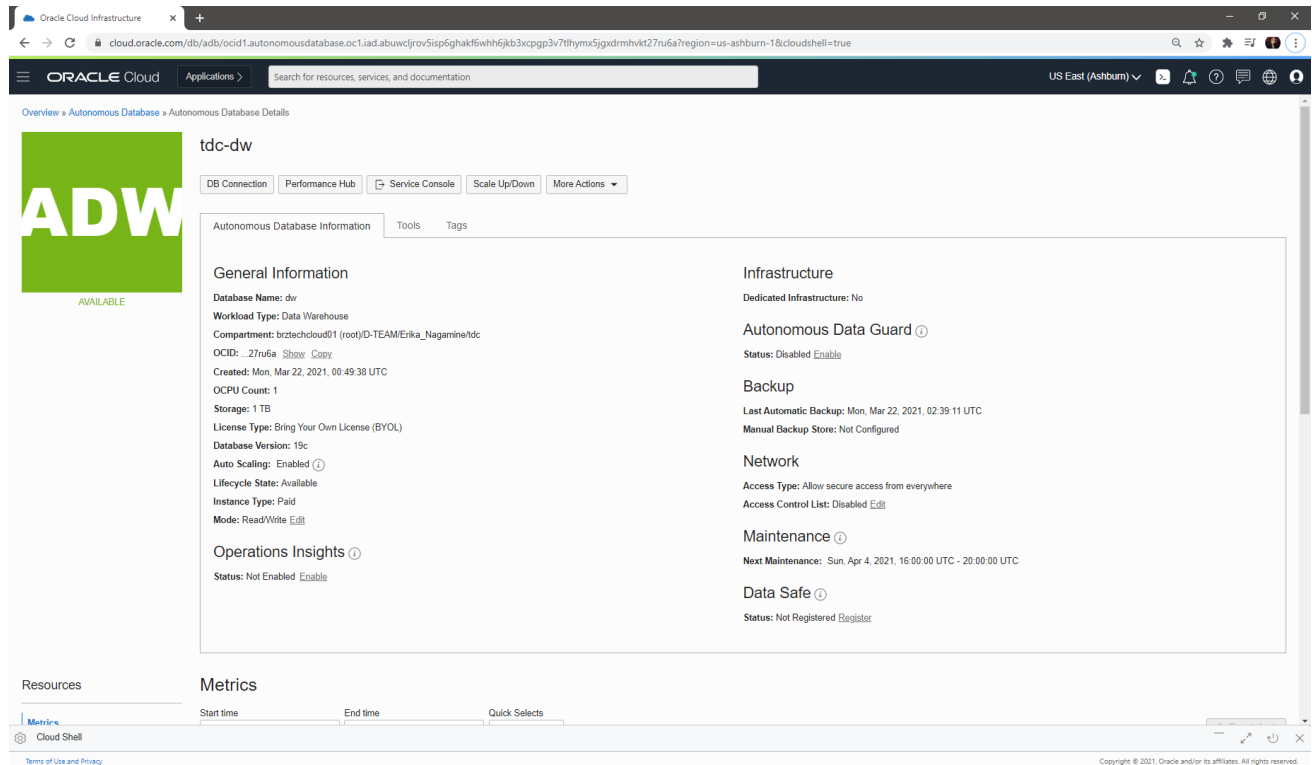
Preencha com as seguintes informações (note que é possível provisionar o autonomous como Always free:

- Compartment: se você criou um compartment, selecione o responsável:
- Display name: tdc-dw
- Database name: dw
- Database version: 19c
- Workload type: data warehouse
- User password: coloque uma de sua escolha
- License type: Byol (se você selecionou a opção Always free do autonomous desconsidere essa parte)

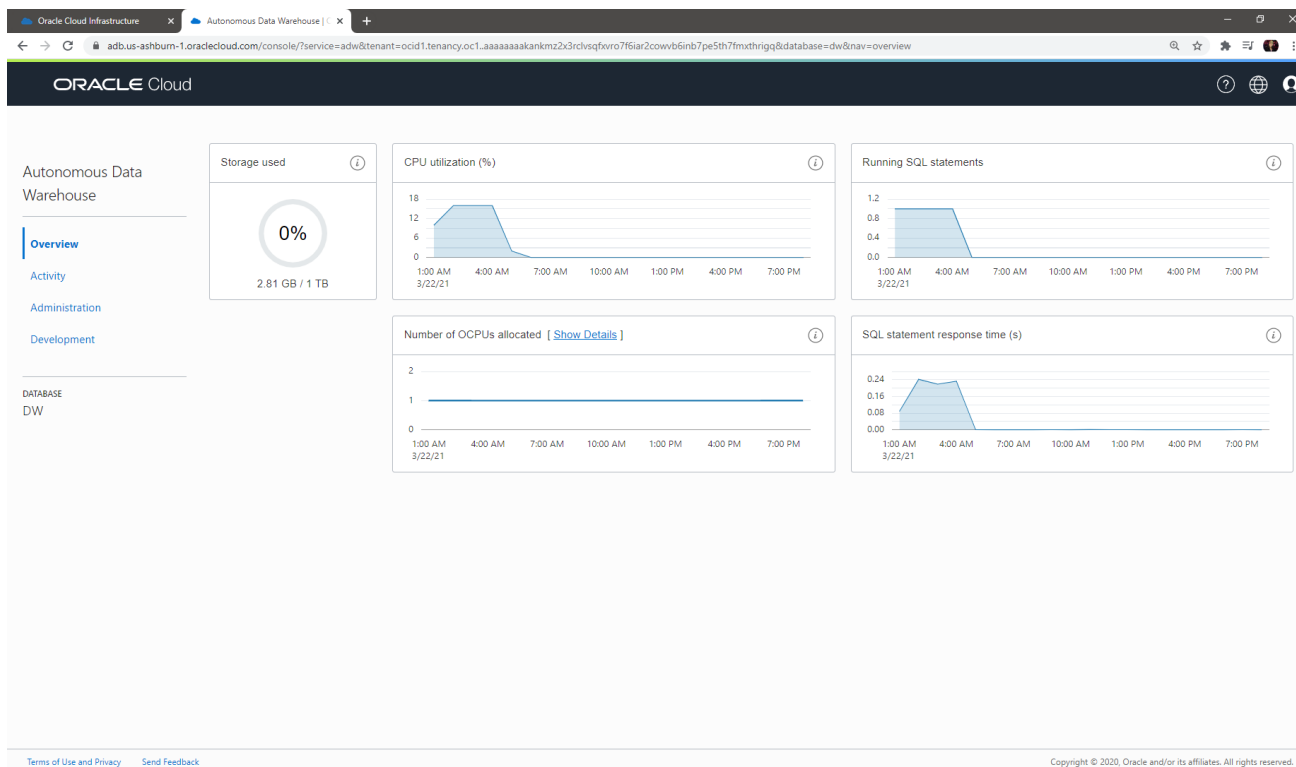
The screenshot displays the Oracle Cloud console interface for an Autonomous Database (ADW) instance. The instance is named 'tdc-dw' and is in an 'AVAILABLE' state. The console shows various tabs for management, including DB Connection, Performance Hub, Service Console, Scale Up/Down, and More Actions. The main content area is divided into sections for General Information, Infrastructure, Autonomous Data Guard, Backup, Network, Maintenance, and Data Safe. The General Information section provides details such as Database Name (dw), Workload Type (Data Warehouse), Compartment (brztechcloud01), OCID, Creation time, OCPU Count, Storage, License Type (Bring Your Own License (BYOL)), Database Version (19c), Auto Scaling (Enabled), Lifecycle State (Available), Instance Type (Paid), and Mode (Read/Write). The Infrastructure section indicates that Dedicated Infrastructure is not used. The Autonomous Data Guard section shows the status as Disabled. The Backup section shows the last automatic backup time. The Network section shows the access type as Allow secure access from everywhere. The Maintenance section shows the next maintenance window. The Data Safe section shows the status as Not Registered.

3. Configurando seu ambiente

Agora que o stack está provisionado, vamos integrar os recursos. O primeiro passo é validar sua conexão via API para o Autonomous database e criar uma collection. Vamos fazer isso tudo através de linha de comando no cloud shell. Vamos pegar a informação necessária para este step, vá no autonomous criado e clique em service console:



Dentro da service console, navegue em development:



Dentro de development, copie a URL do RESTful services and SODA:

Agora substitua as variáveis abaixo e execute no seu


```
export ORDS_BASE_URL=< URL do SEU ADW – Restful and SODA>
export ORDS_USR='<usuario>:<senha>'
```

```
curl -X PUT -u $ORDS_USR -H "Content-Type: application/json"
$ORDS_BASE_URL/admin/soda/latest/regionsnumbers

curl -u $ORDS_USR -H "Content-Type: application/json" $ORDS_BASE_URL/admin/soda/latest/

curl -X POST -u $ORDS_USR -H "Content-Type: application/json" --data '{}'
$ORDS_BASE_URL/admin/soda/latest/regionsnumbers?action=query

curl -X POST -u $ORDS_USR -H "Content-Type: application/json"
$ORDS_BASE_URL/admin/soda/latest/regionsnumbers?action=truncate
```


ORACLE Cloud

Applications >

Search for resources, services, and documentation

Overview » Autonomous Database » Autonomous Database Details



AVAILABLE

tdc-dw

DB Connection
Performance Hub
Service Console
Scale Up/Down
More Actions

Autonomous Database Information Tools Tags

General Information

Database Name: dw
Workload Type: Data Warehouse
Compartment: brztecloud01 (root)/D-TEAM/Erika_Nagamine/tdc
OCID: ...27ru6a [Show](#) [Copy](#)
Created: Mon, Mar 22, 2021, 00:49:38 UTC
OCPU Count: 1
Storage: 1 TB
License Type: Bring Your Own License (BYOL)
Database Version: 19c
Auto Scaling: Enabled ⓘ
Lifecycle State: Available
Instance Type: Paid
Mode: Read/Write [Edit](#)

Operations Insights ⓘ

Cloud Shell

```
erika_naga@cloudshell:~ (us-ashburn-1) $ curl -X POST -u $ORDS_USR -H "Content-Type: application/json" --data '{}' $ORDS_BASE_URL/admin/soda/latest/regionsnumbers?action=query
{"items":[{"id":"1A4CD4FB9967442991CF8044BEAFE137","etag":"7F497F8A397D48DE91DCFB04468F1C82","lastModified":"2021-03-22T02:47:19.756899000Z","created":"2021-03-22T02:47:19.756899000Z","links":{"region":"AWER","col1":"13","col2":"14","col3":"15"}}, {"id":"29ED908437BB4938A5959DE66E97BE1E","etag":"3FDF1B72F47448DD81A2BF71A44C8C17","lastModified":"2021-03-22T02:41:15.028897000Z","snubers/29ED908437BB4938A5959DE66E97BE1E"}], "value":{"region":"AWER","col1":"1","col2":"2","col3":"3"}}, {"id":"4BA395CD6691409FBF765CF86463C931","etag":"95C88D6C9F74407A8E1D6CF82ED89EAB","loudapps.com:443/ords/admin/soda/latest/regionsnumbers/4BA395CD6691409FBF765CF86463C931"}], "value":{"region":"AWER","col1":"22","col2":"23","col3":"24"}}, {"id":"614D378FAE014058881C818126A0D1"
```

Mediante a esse funcionamento, o próximo passo é configurar a parametrização básica da sua função. Para isso vá em developer services -> functions:

The screenshot shows the Oracle Cloud Functions console. The left sidebar has a 'Developer Services' section with links to 'Kubernetes Clusters', 'Container Registry', 'Functions', and 'API Management'. Below this is a 'List Scope' section with a 'Compartment' dropdown set to 'tdc' and a 'Tag Filters' section. The main content area is titled 'Applications in tdc Compartment' and features a 'Create Application' button and a table with one application: 'pipeline', which is 'ACTIVE' and was created on 'Mon, Mar 22, 2021, 00:58:01 UTC'. The bottom of the page shows a 'Cloud Shell' terminal window.

Clique na aplicação criada:

The screenshot shows the details of the 'pipeline' application. The left sidebar has a 'Resources' section with links to 'Getting Started', 'Functions', 'Metrics', 'Network', 'Configuration', and 'Logs'. The main content area shows the application's details, including a large green 'A' icon with 'ACTIVE' below it. The 'Application Information' tab is selected, showing the OCID, logging policy, and creation details. The 'Functions' section below shows a table with two functions: 'oci-load-file-int-o-adv-py' and 'func', both created on Mon, Mar 22, 2021, 02:37:37 UTC and Mon, Mar 22, 2021, 02:04:20 UTC respectively. The bottom of the page shows a 'Cloud Shell' terminal window.

Vá em configuração e sete as seguintes parametrizações:

The screenshot shows the Oracle Cloud console interface for an Autonomous Data Warehouse (ADW) pipeline. The top navigation bar includes the Oracle Cloud logo, a search bar, and the region 'US East (Ashburn)'. The main content area is titled 'pipeline' and features a green circular icon with a white 'A' and the status 'ACTIVE'. Below the icon are buttons for 'Move Application', 'Add Tags', and 'Delete'. The 'Application Information' tab is selected, displaying the OCID, Logging Policy, and Last updated timestamp. The 'Configuration' tab is also visible, showing a table of key-value pairs for the pipeline configuration.

| Key | Value |
|------------------|---|
| dbpwd | Senha do seu ADW |
| ordsbaseurl | URL do Restful Services and SODA (ORDS) |
| dbschema | Usuário |
| namespace | Namespace: obtenha essa info no bucket |
| processed_bucket | processed-bucket |
| bucket_name | input-bucket |
| function | oci-load-file-into-adw-python |
| dbuser | usuário do bd |

Pronto! Agora você está pronto para testar seu pipeline! 😊

4. Testando seu pipeline de dados

Agora que o stack está todo configurado, é só colocar um arquivo no bucket input-bucket. Os arquivos de exemplo podem ser baixados do github:

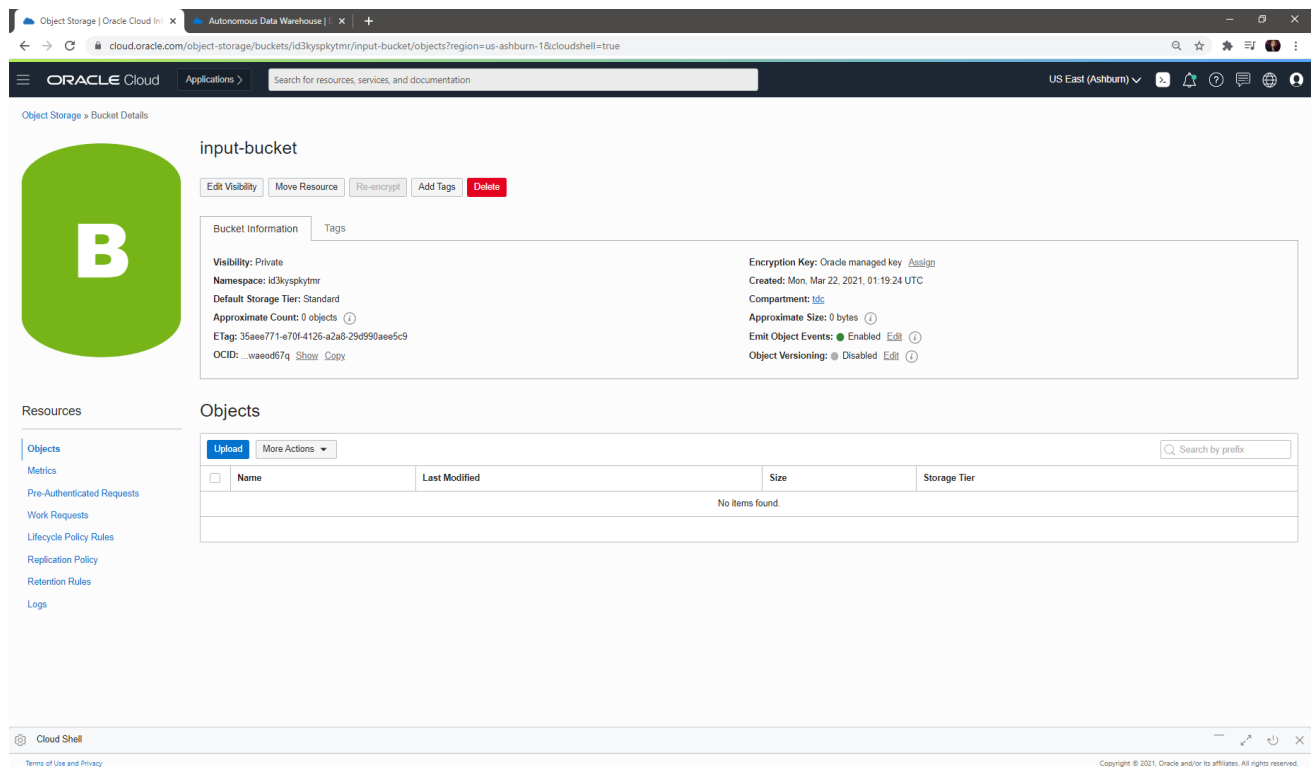
http://

The screenshot displays the Oracle Cloud console interface. At the top, the navigation bar shows 'Object Storage | Oracle Cloud Infrastructure' and 'Autonomous Data Warehouse'. The main content area is titled 'input-bucket' and includes a green bucket icon with a white 'B'. Below the icon, there are tabs for 'Bucket Information' and 'Tags'. The 'Bucket Information' tab is active, showing details such as 'Visibility: Private', 'Namespace: id3kyspkytmr', 'Default Storage Tier: Standard', 'Approximate Count: 0 objects', 'ETag: 35aee771-e70f-4126-a2a8-29d990aee5c9', and 'OCID: ...waed67q'. On the right side of the bucket information, there are fields for 'Encryption Key: Oracle managed key', 'Created: Mon, Mar 22, 2021, 01:19:24 UTC', 'Compartment: idc', 'Approximate Size: 0 bytes', 'Emit Object Events: Enabled', and 'Object Versioning: Disabled'. Below the bucket information, there is a section for 'Objects' with a table that has columns for 'Name', 'Last Modified', 'Size', and 'Storage Tier'. The table is currently empty, showing 'No items found.' On the left side of the console, there is a 'Resources' sidebar with links to 'Objects', 'Metrics', 'Pre-Authenticated Requests', 'Work Requests', 'Lifecycle Policy Rules', 'Replication Policy', 'Retention Rules', and 'Logs'. At the bottom of the console, there is a 'Cloud Shell' section.

The screenshot shows the Oracle Cloud console interface. On the left, the 'input-bucket' details are visible, including its visibility (Private), namespace (id3kyspytmr), and default storage tier (Standard). The main area displays the 'Upload Objects' dialog. It includes a search bar for object names, a storage tier dropdown set to 'Standard', and a file selection area. A file named 'file1.csv' (72 bytes) is selected. Below the selection area, there is a summary of the upload: '1 file, 72 bytes total'. At the bottom of the dialog, there are 'Upload' and 'Cancel' buttons.

This screenshot shows the Oracle Cloud console after the upload process. The 'input-bucket' details are still visible on the left. The main area now displays the 'Objects' list. The list has columns for 'Name', 'Last Modified', 'Size', and 'Storage Tier'. A single object, 'file1.csv', is listed with a last modified date of 'Mon, Mar 22, 2021, 20:53:49 UTC' and a size of '72 bytes'. The storage tier is 'Standard'. The 'Upload' button is still present at the top of the list.

Aguarde um momento até o processo ser concluído. Quando ele for, veja que o arquivo não estará mais presente no bucket input-bucket – o tempo do processo depende do tamanho do arquivo (se for os de exemplo, note que será menos de 1 min):



Object Storage | Oracle Cloud Infrastructure | Autonomous Data Warehouse |

cloud.oracle.com/object-storage/buckets/id3kysplytmr/input-bucket/objects?region=us-ashburn-1&cloudshell=true

ORACLE Cloud Applications > Search for resources, services, and documentation US East (Ashburn)

Object Storage > Bucket Details

input-bucket

Edit Visibility Move Resource Re-encrypt Add Tags Delete

Bucket Information Tags

Visibility: Private
Namespace: id3kysplytmr
Default Storage Tier: Standard
Approximate Count: 0 objects
ETag: 35aee771-e70f-4126-a2a8-29d990aee5c9
OCID: ...waed67q Show Copy

Encryption Key: Oracle managed key Assign
Created: Mon, Mar 22, 2021, 01:19:24 UTC
Compartment: [id3](#)
Approximate Size: 0 bytes
Emit Object Events: Enabled Edit
Object Versioning: Disabled Edit

Resources

- Objects
- Metrics
- Pre-Authenticated Requests
- Work Requests
- Lifecycle Policy Rules
- Replication Policy
- Retention Rules
- Logs

Objects

Upload More Actions Search by prefix

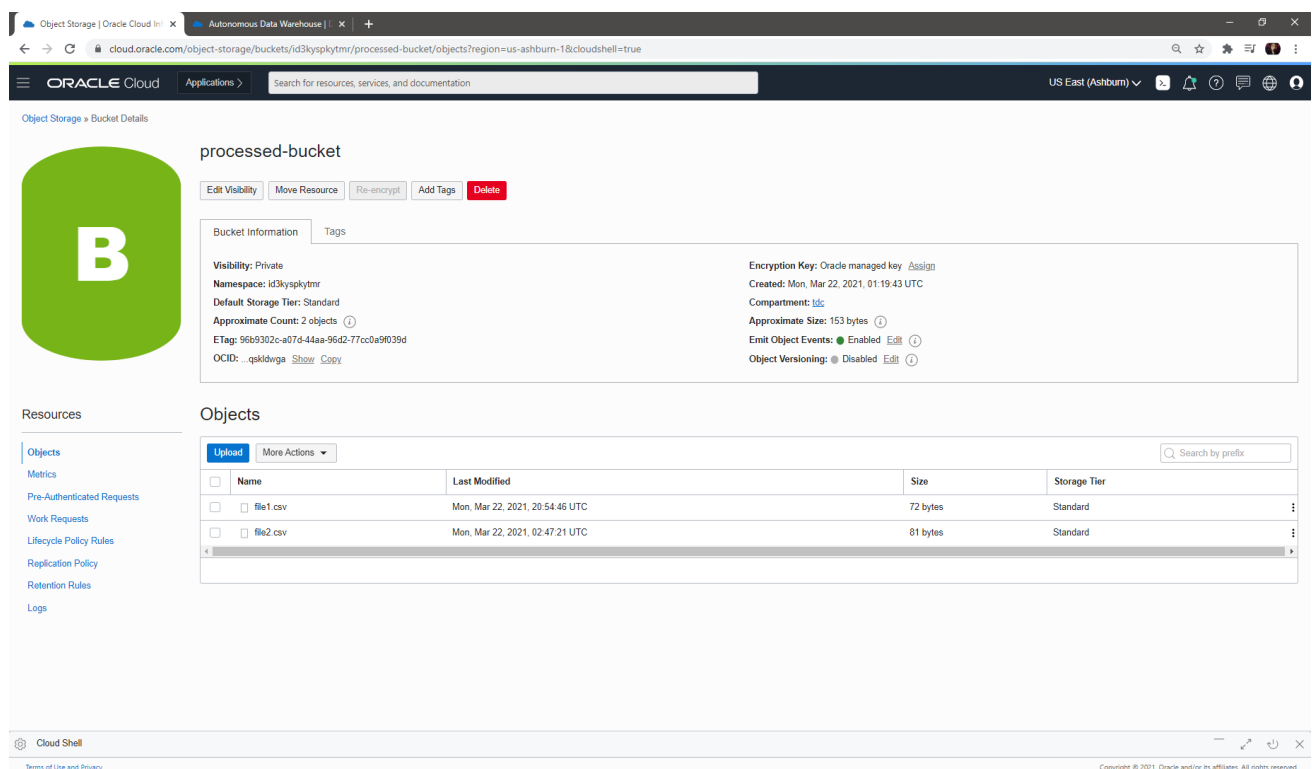
| <input type="checkbox"/> | Name | Last Modified | Size | Storage Tier |
|--------------------------|------|---------------|------|--------------|
| No items found. | | | | |

Cloud Shell

Terms of Use and Privacy

Copyright © 2021, Oracle and/or its affiliates. All rights reserved.

Ao final do processo você terá o arquivo dentro do bucket processed bucket:



Object Storage | Oracle Cloud Infrastructure | Autonomous Data Warehouse |

cloud.oracle.com/object-storage/buckets/id3kysplytmr/processed-bucket/objects?region=us-ashburn-1&cloudshell=true

ORACLE Cloud Applications > Search for resources, services, and documentation US East (Ashburn)

Object Storage > Bucket Details

processed-bucket

Edit Visibility Move Resource Re-encrypt Add Tags Delete

Bucket Information Tags

Visibility: Private
Namespace: id3kysplytmr
Default Storage Tier: Standard
Approximate Count: 2 objects
ETag: 96b9302c-a07d-44aa-96d2-77c0a9f039d
OCID: ...qskldwga Show Copy

Encryption Key: Oracle managed key Assign
Created: Mon, Mar 22, 2021, 01:19:43 UTC
Compartment: [id3](#)
Approximate Size: 153 bytes
Emit Object Events: Enabled Edit
Object Versioning: Disabled Edit

Resources

- Objects
- Metrics
- Pre-Authenticated Requests
- Work Requests
- Lifecycle Policy Rules
- Replication Policy
- Retention Rules
- Logs

Objects

Upload More Actions Search by prefix

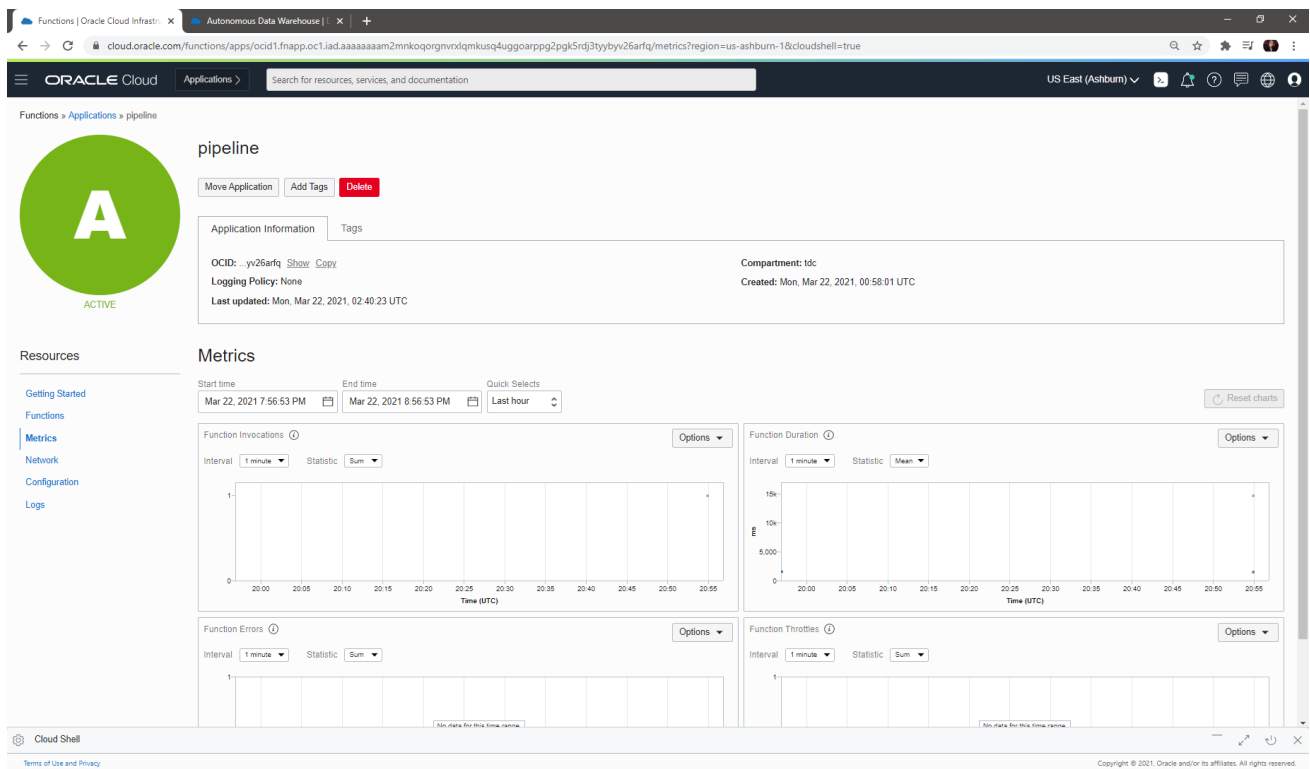
| <input type="checkbox"/> | Name | Last Modified | Size | Storage Tier |
|--------------------------|-----------|---------------------------------|----------|--------------|
| <input type="checkbox"/> | file1.csv | Mon, Mar 22, 2021, 20:54:46 UTC | 72 bytes | Standard |
| <input type="checkbox"/> | file2.csv | Mon, Mar 22, 2021, 02:47:21 UTC | 81 bytes | Standard |

Cloud Shell

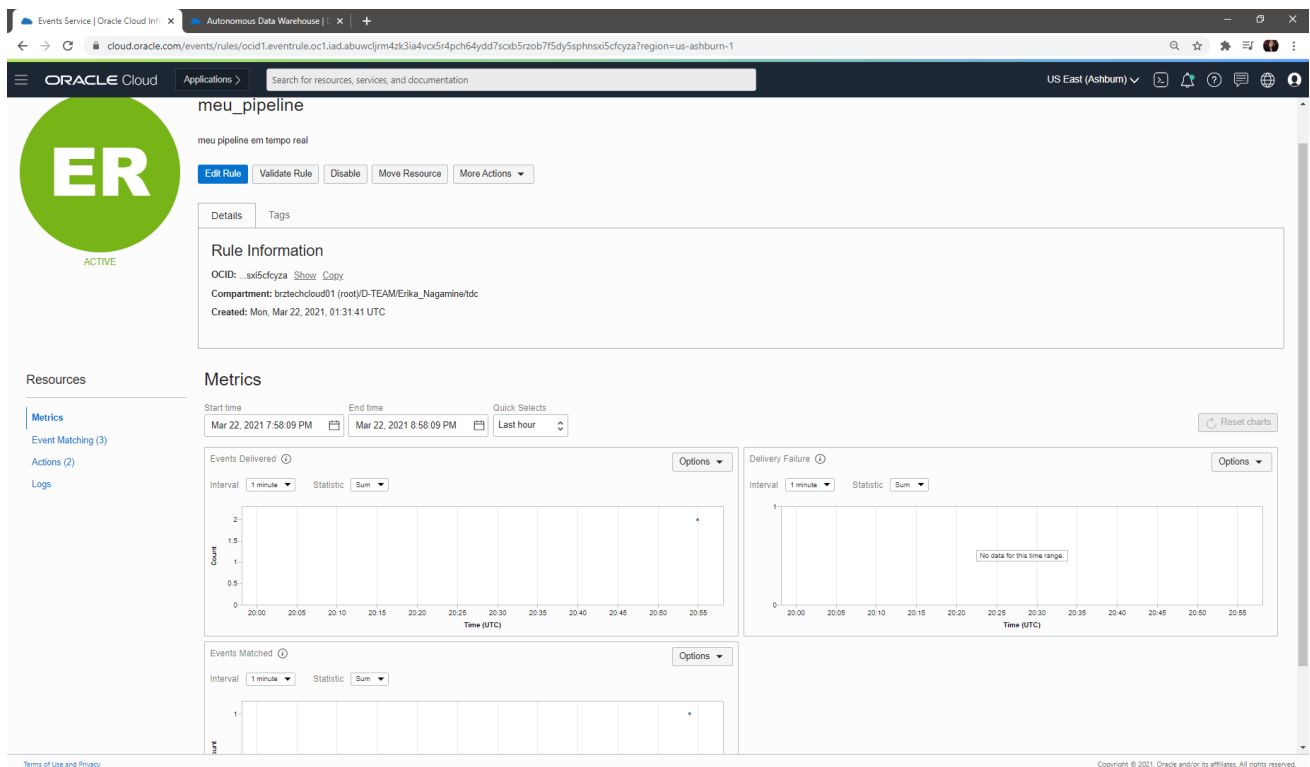
Terms of Use and Privacy

Copyright © 2021, Oracle and/or its affiliates. All rights reserved.

Note que seu processo estará nas métricas da função:



E nas métricas do evento:



Você também pode fazer a validação dentro do seu dw, para isso vá no ADW:

The screenshot displays the Oracle Cloud Infrastructure console for an Autonomous Data Warehouse (ADW) instance named 'tdc-dw'. The instance is in an 'AVAILABLE' state. The console provides a comprehensive overview of the database's configuration and status across several sections:

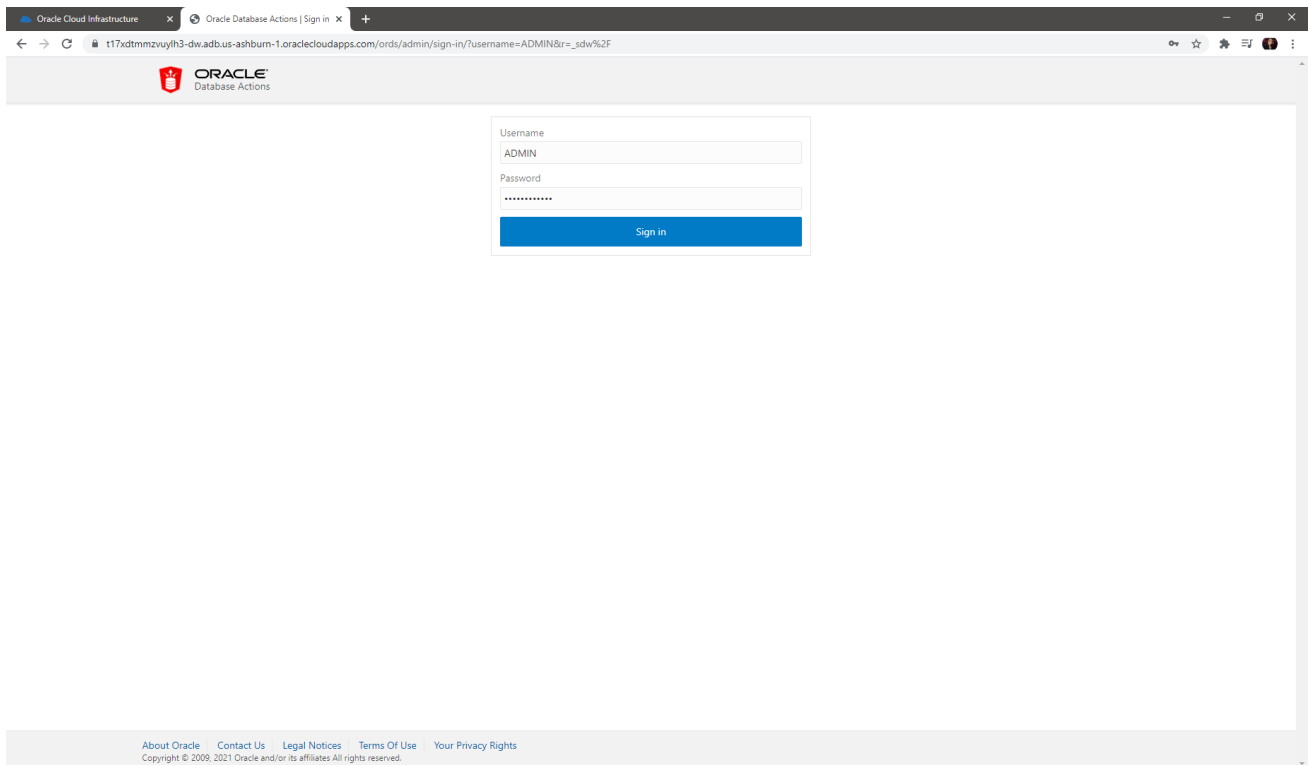
- General Information:** Includes details such as Database Name (dw), Workload Type (Data Warehouse), Compartment (brztechcloud01), OCID, creation date (Mar 22, 2021), OCPU Count (1), Storage (1 TB), License Type (Bring Your Own License), Database Version (19c), Auto Scaling (Enabled), Lifecycle State (Available), Instance Type (Paid), and Mode (Read/Write).
- Infrastructure:** Shows Dedicated Infrastructure as 'No'.
- Autonomous Data Guard:** Status is 'Disabled' with an 'Enable' link.
- Backup:** Last Automatic Backup occurred on Mar 22, 2021, at 02:39:11 UTC. Manual Backup Store is 'Not Configured'.
- Network:** Access Type is 'Allow secure access from everywhere' and Access Control List is 'Disabled'.
- Maintenance:** Next Maintenance is scheduled for Sun, Apr 4, 2021, from 16:00:00 UTC to 20:00:00 UTC.
- Data Safe:** Status is 'Not Registered' with a 'Register' link.
- Operations Insights:** Status is 'Not Enabled' with an 'Enable' link.

At the bottom, there are tabs for 'Resources' (Metrics, Backups) and a 'Metrics' section with filters for Start time, End time, and Quick Selects. A 'Reset charts' button is also present.

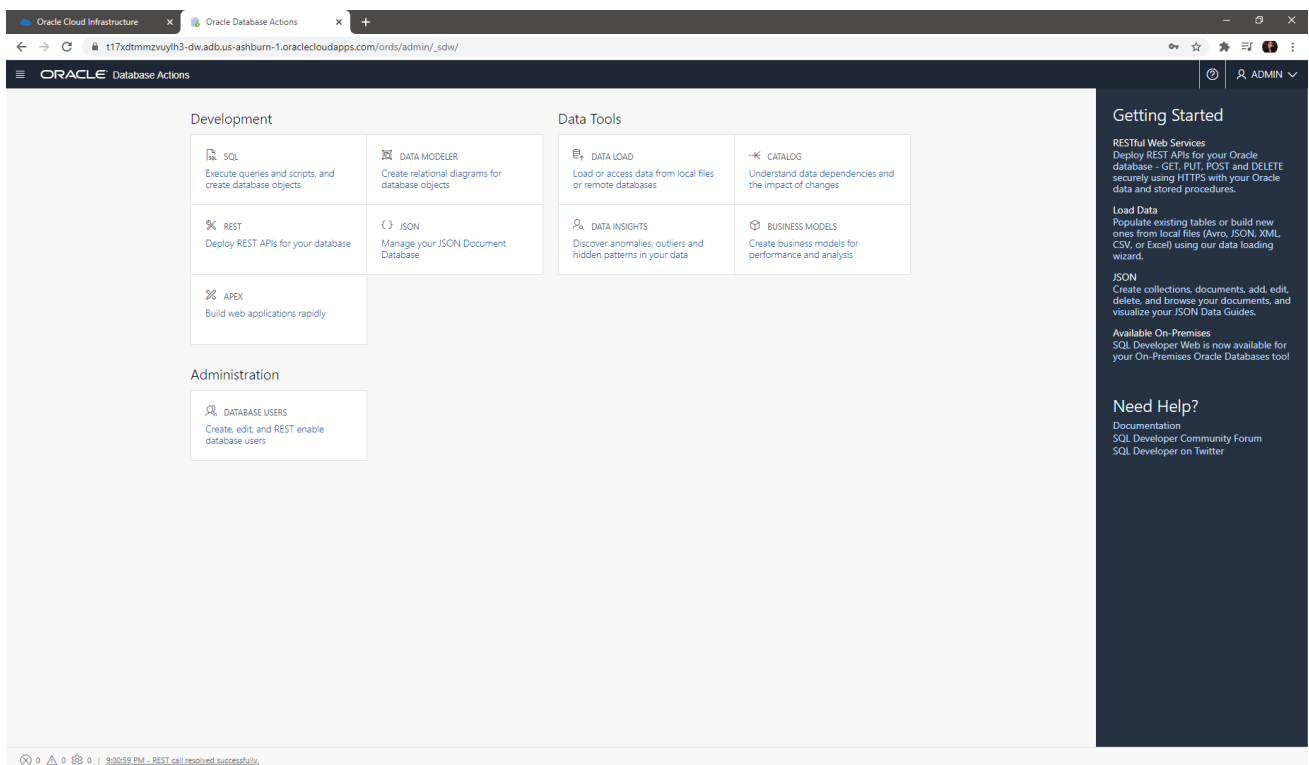
Clique em tools -> database actions:

The screenshot shows the Oracle Database Actions console. It features a login interface with a 'Username' input field and a blue 'Next' button. The Oracle logo and 'Database Actions' text are visible at the top. The footer contains links for 'About Oracle', 'Contact Us', 'Legal Notices', 'Terms Of Use', and 'Your Privacy Rights', along with a copyright notice for 2009-2021.

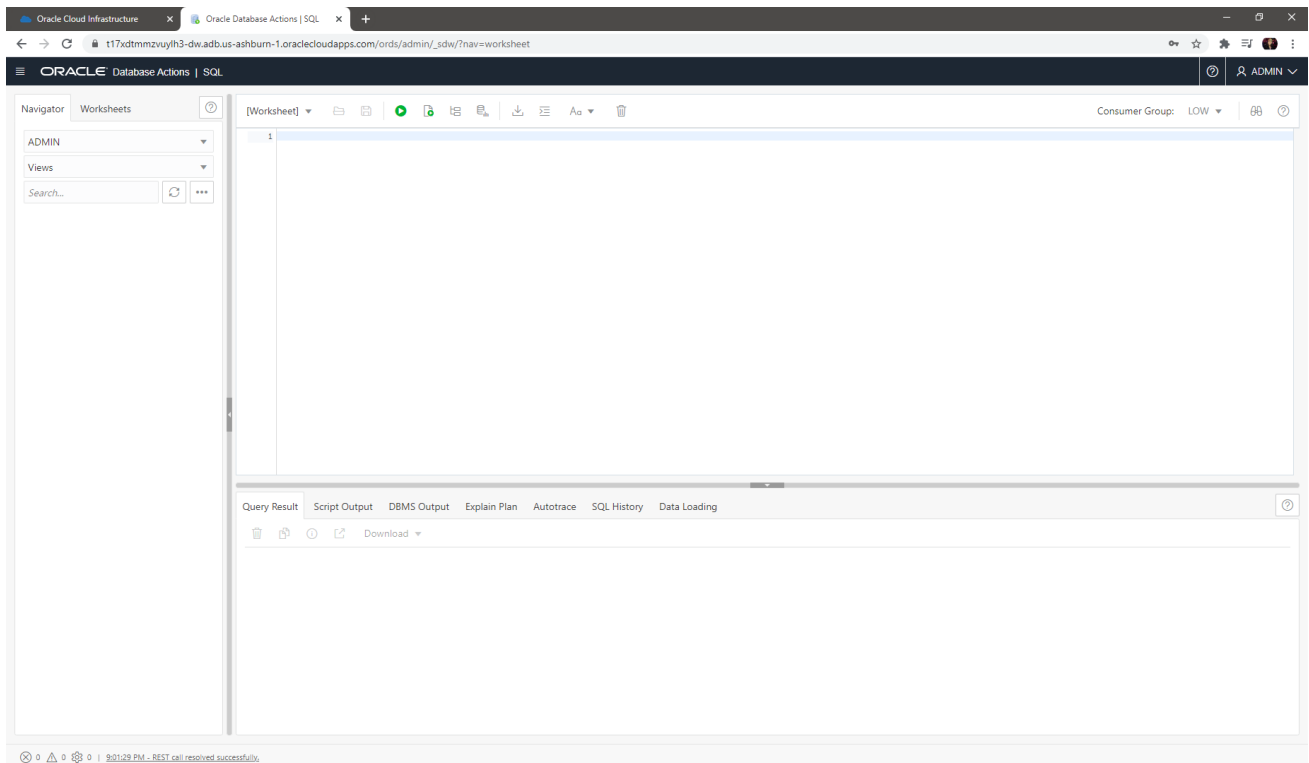
Insira seu usuário e senha:



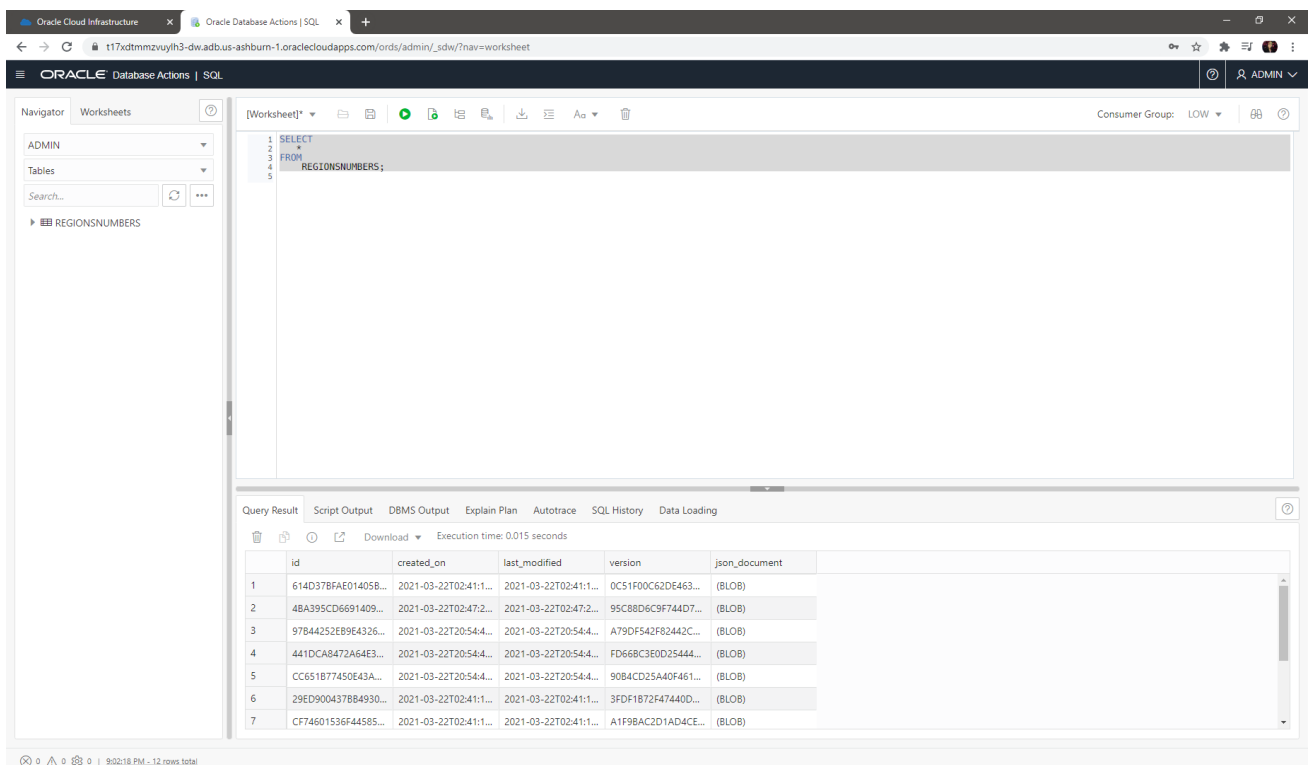
Clique em SQL:



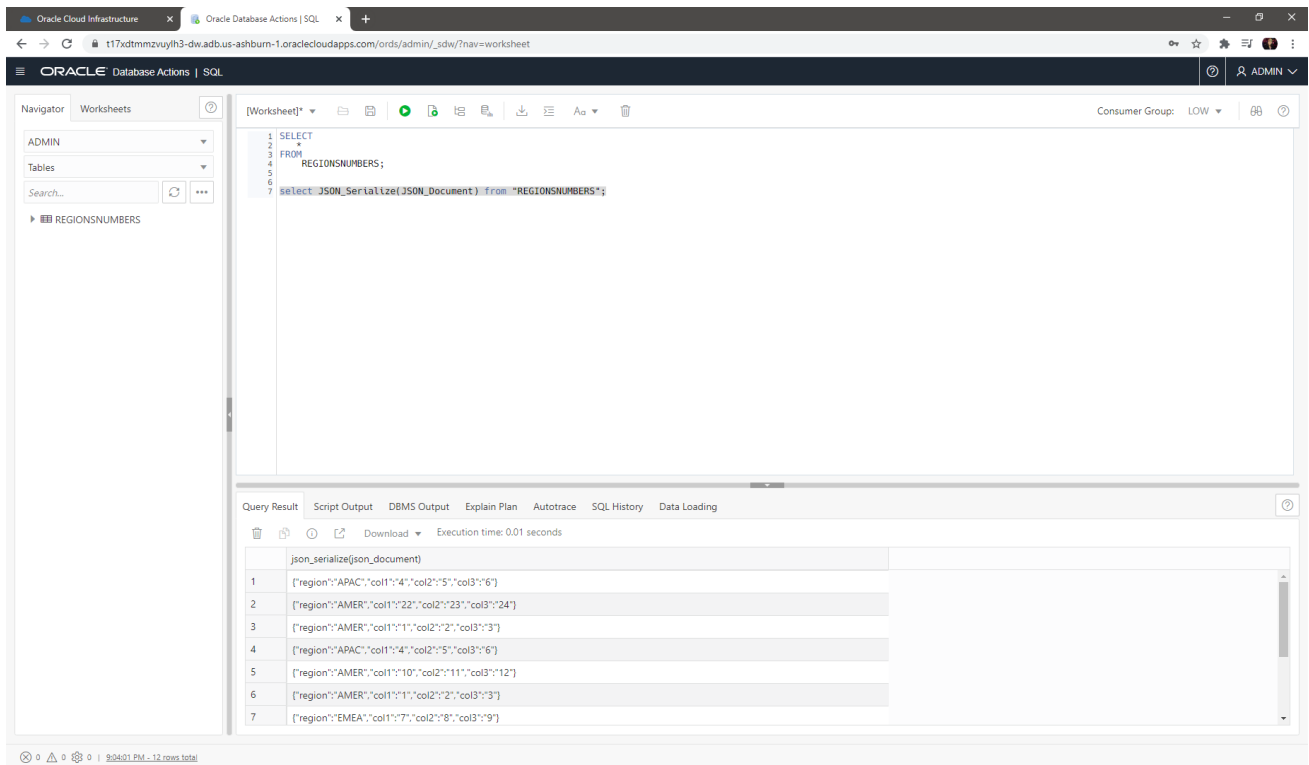
Na tela de SQL, faça a query no objeto criado:



Ao realizar q query note que ela aparece como objeto json:



Veja que é possível melhorar a visualização utilizando os recursos json:



The screenshot displays the Oracle Cloud Infrastructure (OCI) console interface for a database action. The top navigation bar shows the OCI logo and the current page is "Oracle Database Actions | SQL". The breadcrumb trail indicates the path: "t17xdtmmzvuyh3-dw.adb.us-ashburn-1.oraclecloudapps.com/ords/admin/_sdw/?nav=worksheet".

The main interface is divided into two main sections: a left sidebar and a central workspace.

Left Sidebar:

- Navigator:** Shows the "ADMIN" tab selected. Under "Tables", the "REGIONSNUMBERS" table is listed.
- Worksheets:** A tab labeled "[Worksheet]*" is active.

Central Workspace:

The SQL editor contains the following query:

```
1 SELECT
2 *
3 FROM
4   REGIONSNUMBERS;
5
6 select JSON_Serialize(JSON_Document) from "REGIONSNUMBERS";
7
```

Below the editor, the "Query Result" tab is selected, showing the output of the query. The execution time is 0.01 seconds. The result is a table with 7 rows of JSON data:

| | json_serialize(json_document) |
|---|--|
| 1 | ["region": "APAC", "col1": "4", "col2": "5", "col3": "6"] |
| 2 | ["region": "AMER", "col1": "22", "col2": "23", "col3": "24"] |
| 3 | ["region": "AMER", "col1": "1", "col2": "2", "col3": "3"] |
| 4 | ["region": "APAC", "col1": "4", "col2": "5", "col3": "6"] |
| 5 | ["region": "AMER", "col1": "10", "col2": "11", "col3": "12"] |
| 6 | ["region": "AMER", "col1": "1", "col2": "2", "col3": "3"] |
| 7 | ["region": "EMEA", "col1": "7", "col2": "8", "col3": "9"] |

At the bottom of the console, a status bar indicates the time: "9:24:01 PM - 12 rows total".

5. Referências

Algumas documentações importantes:

- Functions: https://www.oracle.com/webfolder/technetwork/tutorials/infographics/oci_functions_cloudshell_quickview/functions_quickview_top/functions_quickview/index.html#
- OCI Examples: <https://github.com/oracle/oracle-functions-samples/tree/master/samples/oci-load-file-into-adw-python>
- Archicture Reference: <https://docs.oracle.com/en/solutions/serverless-dataload-adw/index.html#GUID-36A94B9D-0B06-4A4A-81B0-A292392181BD>