

Data Engineering

Filipe Nascimento
fgvn@cesar.school





C . E . S . A . R

Pessoas impulsionando inovação.
Inovação impulsionando negócios.

Everton Dias
etgdb@cesar.org.br

Janaína Branco
jcb@cesar.org.br



AWS BOT03

A computação em cloud traz diversas facilidades e possibilidades ao nosso cotidiano, em aulas passadas estudamos brevemente o S3 - serviço de armazenamento de objetos da AWS. Hoje aprenderemos a manipular buckets através da biblioteca boto3.

CONCEITOS

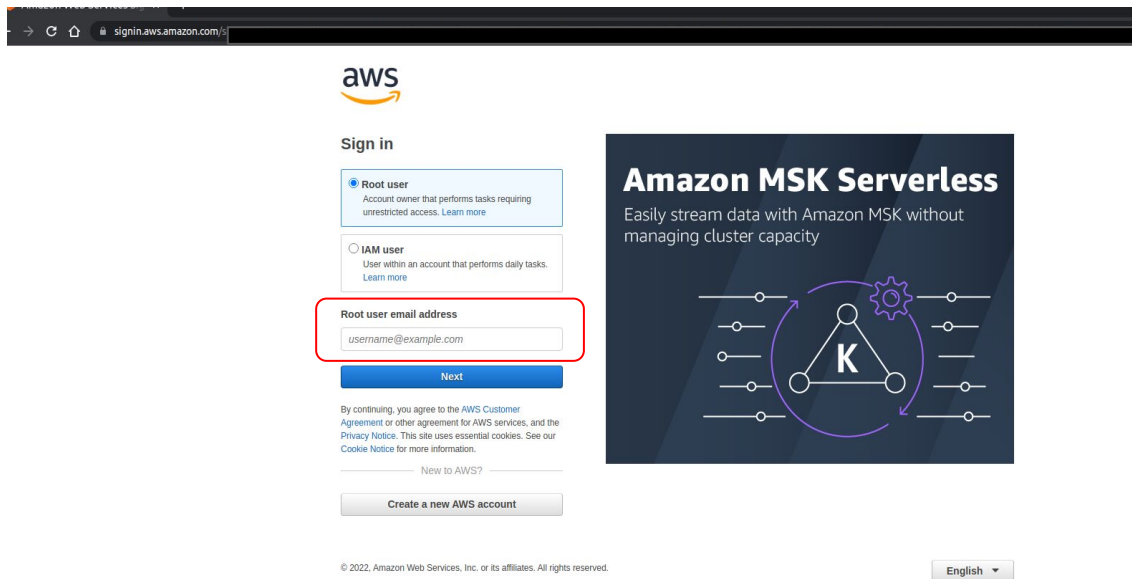
Usuário IAM

Antes de iniciarmos nossa atividade, é importante que seja criado um usuário IAM e credenciais de acesso programático.

1. Acesse o seu console da AWS console.aws.amazon.com;
2. Utilize seu e-mail e senha com o qual se inscrever na aws.amazon.com;
3. Navegue até o serviço IAM - Identity and Access Management;
4. Clique na opção Users / Usuários na coluna do lado esquerdo;
5. Clique na opção no canto superior direito - Add Users / Adicionar Usuários;
6. Coloque um nome para seu usuário e marque a opção de geração de chaves;
7. Selecione as permissões referentes ao S3;
8. Passe pela opção de Adicionar Tags e Review e por fim clique no botão de Create User / Criar Usuário;
9. Na tela seguinte você terá as informações referentes ao ID e Secret das chaves de acesso para realizar o tutorial no notebook disponibilizado.

PASSO 1 - Acesse o seu console da AWS console.aws.amazon.com;

PASSO 2 - Utilize seu e-mail e senha com o qual se inscreveu na aws.amazon.com;



The screenshot shows the AWS Sign in page in a web browser. The address bar displays `signin.aws.amazon.com`. The page features the AWS logo at the top. Below it, the "Sign in" section offers two options: "Root user" (selected) and "IAM user". The "Root user" option is described as "Account owner that performs tasks requiring unrestricted access." Below these options, the "Root user email address" field is highlighted with a red rectangle, containing the text `username@example.com`. A blue "Next" button is positioned below the email field. To the right of the sign-in form is a promotional banner for "Amazon MSK Serverless" with the text "Easily stream data with Amazon MSK without managing cluster capacity" and a diagram of a Kinesis cluster. At the bottom of the page, there is a link for "Create a new AWS account" and a language selector set to "English".

aws

Sign in

☒ **Root user**
Account owner that performs tasks requiring unrestricted access. [Learn more](#)

☐ **IAM user**
User within an account that performs daily tasks. [Learn more](#)

Root user email address

Next

By continuing, you agree to the [AWS Customer Agreement](#) or other agreement for AWS services, and the [Privacy Notice](#). This site uses essential cookies. See our [Cookie Notice](#) for more information.

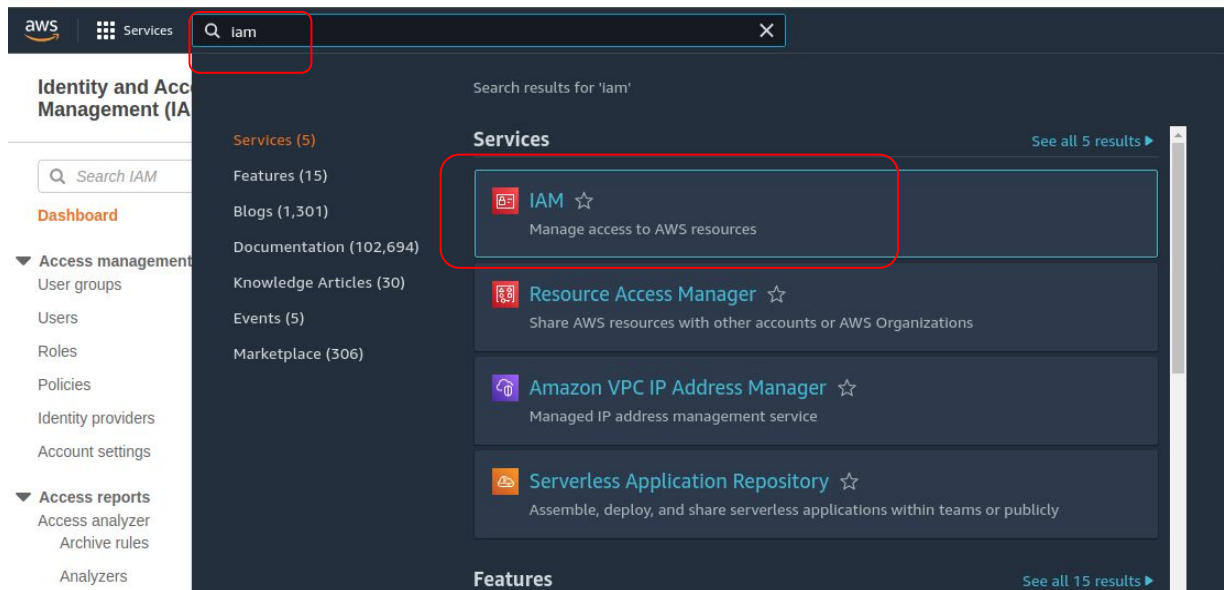
☐ New to AWS?

Create a new AWS account

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

English

PASSO 3 - Navegue até o serviço IAM - Identity and Access Management;



The screenshot shows the AWS IAM console search results for 'iam'. The search bar at the top left contains 'iam'. The left sidebar shows the 'Identity and Access Management (IAM)' section selected. The main content area displays search results for 'iam', including a list of services and features. The 'IAM' service is highlighted with a red box. The 'IAM' service is described as 'Manage access to AWS resources'. Other services listed include 'Resource Access Manager', 'Amazon VPC IP Address Manager', and 'Serverless Application Repository'. The 'Features' section is also visible at the bottom.

aws Services

Identity and Access Management (IAM)

Dashboard

Access management

- User groups
- Users
- Roles
- Policies
- Identity providers
- Account settings

Access reports

- Access analyzer
- Archive rules
- Analyzers

Search results for 'iam'

Services (5)

- Features (15)
- Blogs (1,301)
- Documentation (102,694)
- Knowledge Articles (30)
- Events (5)
- Marketplace (306)

Services

- IAM** ☆
Manage access to AWS resources
- Resource Access Manager** ☆
Share AWS resources with other accounts or AWS Organizations
- Amazon VPC IP Address Manager** ☆
Managed IP address management service
- Serverless Application Repository** ☆
Assemble, deploy, and share serverless applications within teams or publicly

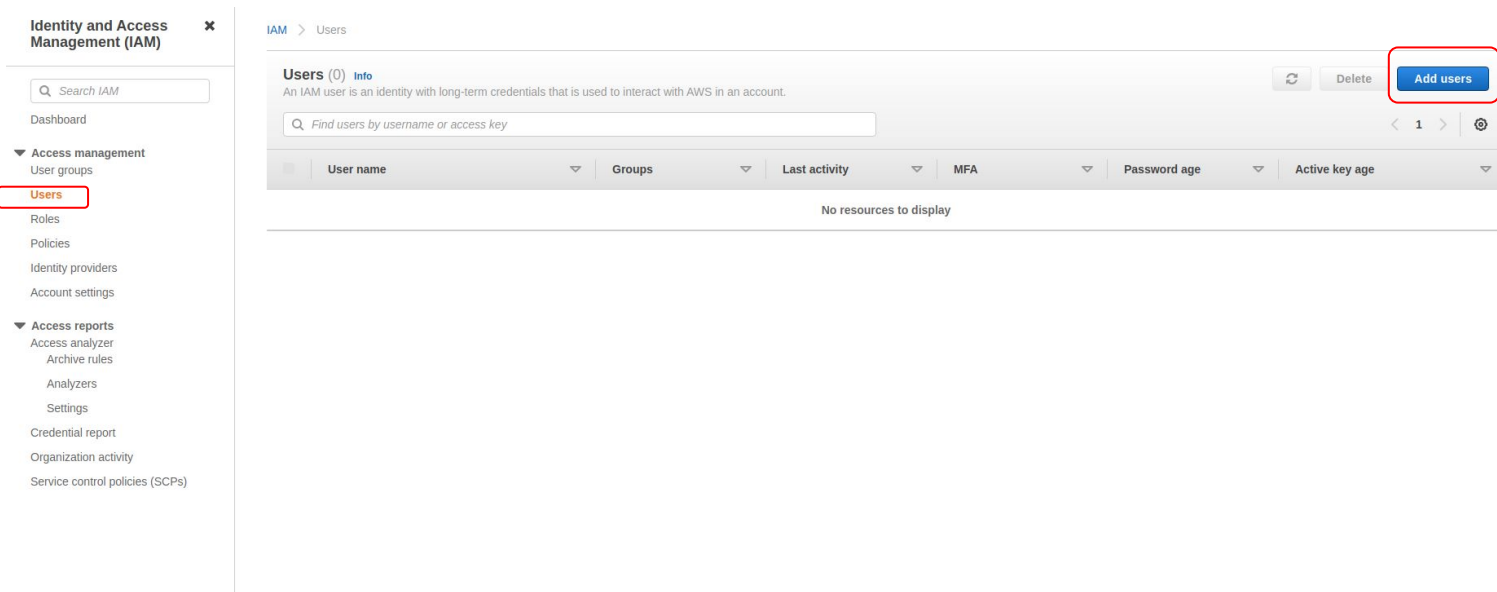
Features

See all 5 results ▶

See all 15 results ▶

PASSO 4 - Clique na opção Users / Usuários na coluna do lado esquerdo;

PASSO 5 - Clique na opção no canto superior direito - Add Users / Adicionar Usuários;



The screenshot displays the AWS IAM console interface. On the left, the 'Identity and Access Management (IAM)' sidebar is visible, with the 'Users' link under 'Access management' highlighted by a red rectangle. The main content area shows the 'Users (0)' page, which includes a search bar and a table with columns for User name, Groups, Last activity, MFA, Password age, and Active key age. The 'Add users' button in the top right corner is highlighted by a red rectangle. The status 'No resources to display' is shown at the bottom of the table area.

PASSO 6 - Coloque um nome para seu usuário e marque a opção de geração de chaves;


logs, docs, and more [Alt+S]

Add user

1 2 3 4 5

Set user details


You can add multiple users at once with the same access type and permissions. [Learn more](#)

 User name*

[Add another user](#)

Select AWS access type

Select how these users will primarily access AWS. If you choose only programmatic access, it does NOT prevent users from accessing the console using an assumed role. Access keys and autogenerated passwords are provided in the last step. [Learn more](#)

 Select AWS credential type* ☒ **Access key - Programmatic access**
Enables an **access key ID** and **secret access key** for the AWS API, CLI, SDK, and other development tools.

☐ **Password - AWS Management Console access**
Enables a **password** that allows users to sign-in to the AWS Management Console.


PASSO 7 - Selecione as permissões referentes ao S3 - AmazonS3FullAccess;


Add user

1 2 3 4 5

▼ Set permissions

 Add user to group

 Copy permissions from existing user

 Attach existing policies directly










Create policy



Filter policies ▼

Q s3

Showing 9 results

	Policy name ▼	Type	Used as
<input type="checkbox"/>	 AmazonDMSRedshiftS3Role	AWS managed	None
<input checked="" type="checkbox"/>	 AmazonS3FullAccess	AWS managed	Permissions policy (2)
<input type="checkbox"/>	 AmazonS3ObjectLambdaExecutionRolePolicy	AWS managed	None
<input type="checkbox"/>	 AmazonS3OutpostsFullAccess	AWS managed	None
<input type="checkbox"/>	 AmazonS3OutpostsReadOnlyAccess	AWS managed	None
<input type="checkbox"/>	 AmazonS3ReadOnlyAccess	AWS managed	None
<input type="checkbox"/>	 AWSBackupServiceRolePolicyForS3Backup	AWS managed	None
<input type="checkbox"/>	 AWSBackupServiceRolePolicyForS3Restore	AWS managed	None
<input type="checkbox"/>	 QuickSightAccessForS3StorageManagementAnalyticsReadOnly	AWS managed	None

► Set permissions boundary

Cancel

Previous

Next: Tags

OBS: Ao término do exercício exclua o usuário criado.

PASSO 8 - Passe pela opção de Adicionar Tags e Review e por fim clique no botão de Create User / Criar Usuário;

Add user

1 2 3 4 5

Review

Review your choices. After you create the user, you can view and download the autogenerated password and access key.

User details

User name	eng_dados_cesar_renner_01
AWS access type	Programmatic access - with an access key
Permissions boundary	Permissions boundary is not set

Permissions summary

The following policies will be attached to the user shown above.

Type	Name
Managed policy	AmazonS3FullAccess

Tags

No tags were added.



Cancel Previous **Create user**

OBS: Ao término do exercício exclua o usuário criado.

PASSO 9 - Na tela seguinte você terá as informações referentes ao ID e Secret das chaves de acesso para realizar o tutorial no notebook disponibilizado.

Add user

1 2 3 4 5



Success

You successfully created the users shown below. You can view and download user security credentials. You can also email users instructions for signing in to the AWS Management Console. This is the last time these credentials will be available to download. However, you can create new credentials at any time.

Users with AWS Management Console access can sign-in at: <https://675686616425.signin.aws.amazon.com/console>



Download .csv

	User	Access key ID	Secret access key
▶	eng_dados_encantech	AKIAZ2UQPRVU4RJ5AXVC	***** Show



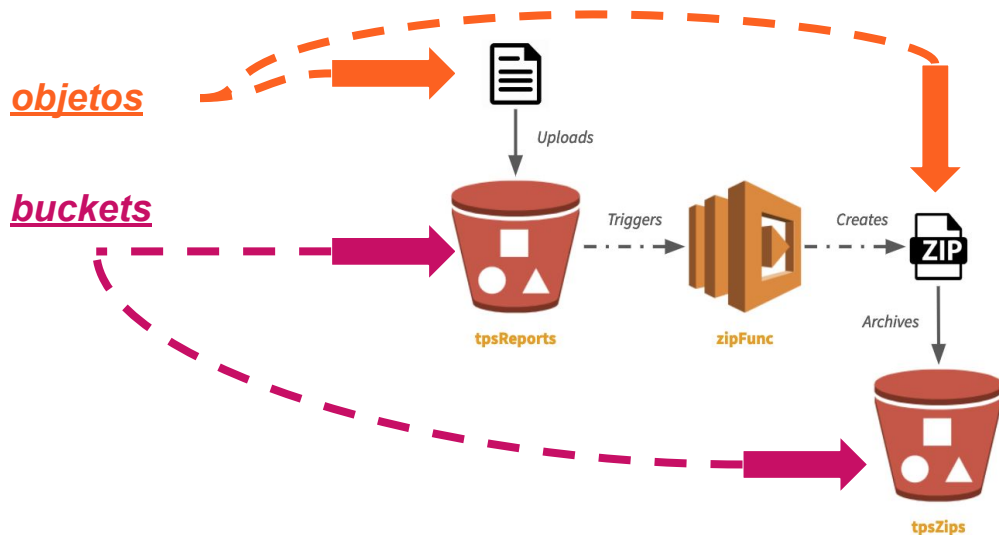
AWS S3

O serviço de storage S3 (Simple Storage Service) nos permite colocar qualquer arquivo na nuvem e torná-lo acessível no mundo todo através de uma URL, gerir recursos de armazenamento na nuvem é um dos pontos chaves das pipelines de dados e do dia a dia do engenheiro de dados.



AWS S3

O S3 possui basicamente dois componentes principais, os buckets e os objetos.



AWS S3

Buckets	Objects
<i>Um bucket possui um 'Name'</i>	<i>Um objeto possui uma 'Key'</i>
<i>'Name' de um bucket é uma String</i>	<i>'Name' de um objeto é o caminho completo desde a raiz do bucket (root)</i>
<i>'Name' de um bucket é único em todo o serviço S3</i>	<i>Uma 'Key' é única dentro do bucket</i>
<i>Contém muitos Objects</i>	<i>Pode existir em apenas um bucket por</i>

AWS BOTO3

Siga os exercícios que estão disponibilizados no link do google colab presente no classroom.

Para mais informações leia a documentação do boto3 neste link:

<https://boto3.amazonaws.com/v1/documentation/api/latest/guide/quickstart.html>

PRÁTICA

Exercício:

Comumente as pipelines realizam a coleta, transporte, transformação e entrega dos dados de uma fonte até seu repositório, ou de um repositório central aos clientes internos (times de análise de dados e etc...).

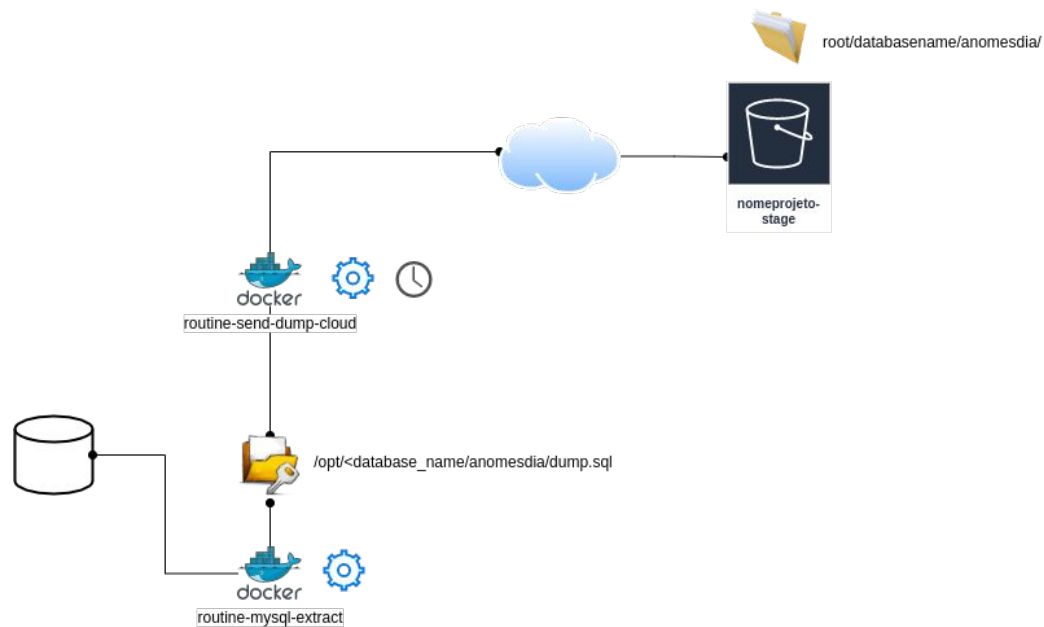
Dentro de tal cenário é bastante comum a divisão dos estágios de coleta, transformação e carga dos dados.

Hoje trabalharemos com o notebook disponibilizado no google colab no link:

<<https://...>>

Construiremos uma automação para criar os buckets referentes aos estágios e automatizar o processo de envio dos dados para a cloud.

Exercício:





ATÉ AMANHÃ!