YouTube Data Analysis

Project Overview

- The requirement from (our simulated) customer:

  ✓ Launching new data-driven campaign
  ✓ Main advertising channel : YouTube
  ✓ Initial question to Answer:
    o "how to categories videos, based on their comments and statistics"
    o "what factors affect how popular a YouTube video will be"

- Why YouTube?

  Top three most visited website (monthly)
    ✓ Google : 92.5 billion
    ✓ YouTube : 34.6 billion
    ✓ Facebook : 25.5 billion

- Goals And Success Criteria

  How my customer will measure success?
    ✓ Data ingestion - Ingest data one ops and incrementally
    ✓ Data Lake - design and build a new Data Lake architecture
    ✓ AWS cloud - AWS as the cloud provider
    ✓ ETL design – extract, transform and load data efficiently
    ✓ Scalability - the data architecture should scale efficiently
    ✓ Reporting - build a business intelligence tier including dashboards

- What is Big Data

  Big data is a term for:
    ✓ Massive data sets with varied and complex structure
    ✓ Which the difficulties of storing and analyzing

- ✓ Visualizing for further Processes or results

- Sources:

  kaggle data collected using YouTube API

  Download data from here: -
  https://www.kaggle.com/datasnaek/youtube-new

- What is on-premise data center

  - ✓ You / Your company has its own hardware
  - ✓ You will need to maintain it example:
    - o Purchase fix and upgrade hardware
    - o Install and maintain operating system and other software

- What is Cloud?

  - ✓ Applications delivered as service over the Internet Example Google platform AWS Azure etc.
  - ✓ Hardware and systems software in data centers that provide those services

- What is AWS?

  - ✓ The leading cloud platform
  - ✓ 200+ services
    - o Compute
    - o Storage and databases
    - o AI and ML
    - o Data Lakes and Analytics
    - o Internet of things

- Step-1: Create an AWS account for your lab

aws

Sign up for AWS

Confirm your identity ⓘ Info

Name ⓘ Info
Choose the name that you want to use for identity verification.

🔵 Shrejal Upadhyay

Primary purpose of account registration
Choose one that best applies to you. If your account is tied to a business, select the one that applies to your business.

Personal use ▼

Ownership type

Individual ▼

✓ Best practices for your AWS account:
- o Always protect your root account
- o Don't use your root account
- o Always unable MFA whenever you can
  - ▪ Example Google authenticator
- o Rotate all keys and passwords periodically
- o Avoid credentials file in shared or even in private computers
- o Follow least privilege principle always

- Step-2: create an IAM user and IAM group for this training

So this is important to get this concerned apply once you create your account what do you need to do you need to go IAM identity access, And there we will create one more account.
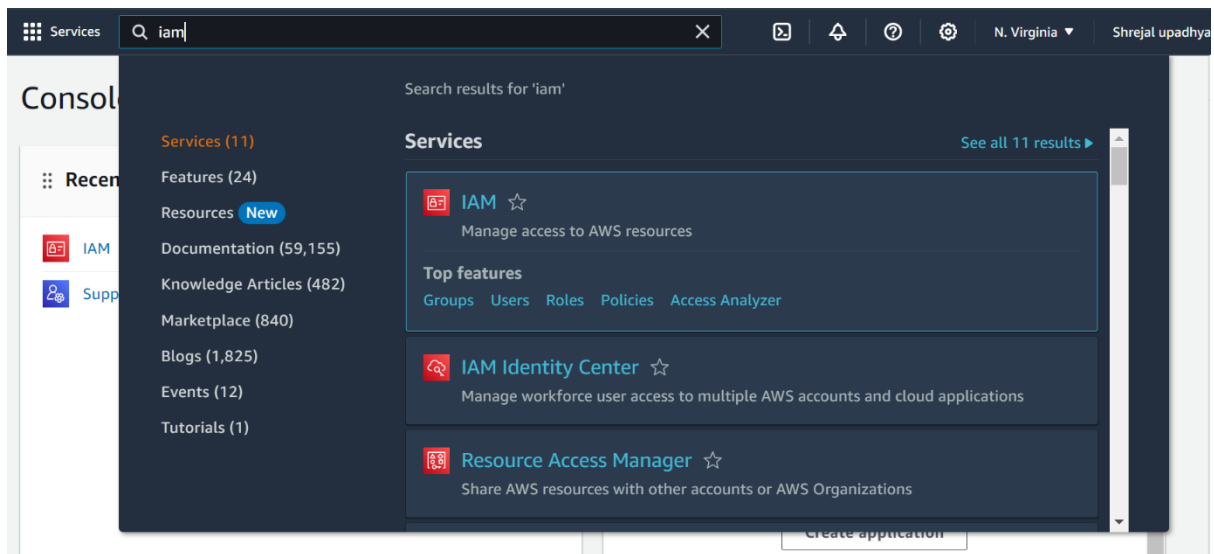
✓ Go to User-> Add User -> Give name for your IAM user name -> Select both checkbox Access key- programmatic access key & Password -> Create Custom password -> keep uncheck Require password reset.

Console password

○ Autogenerated password
   You can view the password after you create the user.

● Custom password
   Enter a custom password for the user.

   [   ]

   • Must be at least 8 characters long
   • Must include at least three of the following mix of character types: uppercase letters (A-Z), lowercase letters (a-z), numbers (0-9), and symbols ! @ # $ % ^ & * ( ) _ + - (hyphen) = [ ] { } | '

☑ Show password

☐ Users must create a new password at next sign-in - Recommended
   Users automatically get the IAMUserChangePassword ↗ policy to allow them to change their own password.

ⓘ If you are creating programmatic access through access keys or service-specific credentials for AWS CodeCommit or Amazon Keyspaces, you can generate them after you create this IAM user. Learn more ↗

Cancel   Next

✓ Click on Attach existing policies-> Check on AdministratorAccess

Step 1
Specify user details

Step 2
Set permissions

Step 3
Review and create

Step 4
Retrieve password

Set permissions

Add user to an existing group or create a new one. Using groups is a best-practice way to manage user's permissions by job functions. Learn more ↗

Permissions options

○ Add user to group
Add user to an existing group, or create a new group. We recommend using groups to manage user permissions by job function.

○ Copy permissions
Copy all group memberships, attached managed policies, and inline policies from an existing user.

● Attach policies directly
Attach a managed policy directly to a user. As a best practice, we recommend attaching policies to a group instead. Then, add the user to the appropriate group.

Permissions policies (1222)

⟳   Create policy ↗

Choose one or more policies to attach to your new user.

✓ Click on Next-> Next->Create User-> After that Download Your Credentials

✓ So once you have created logout from your root account and login with your IAM account, that is look like this->
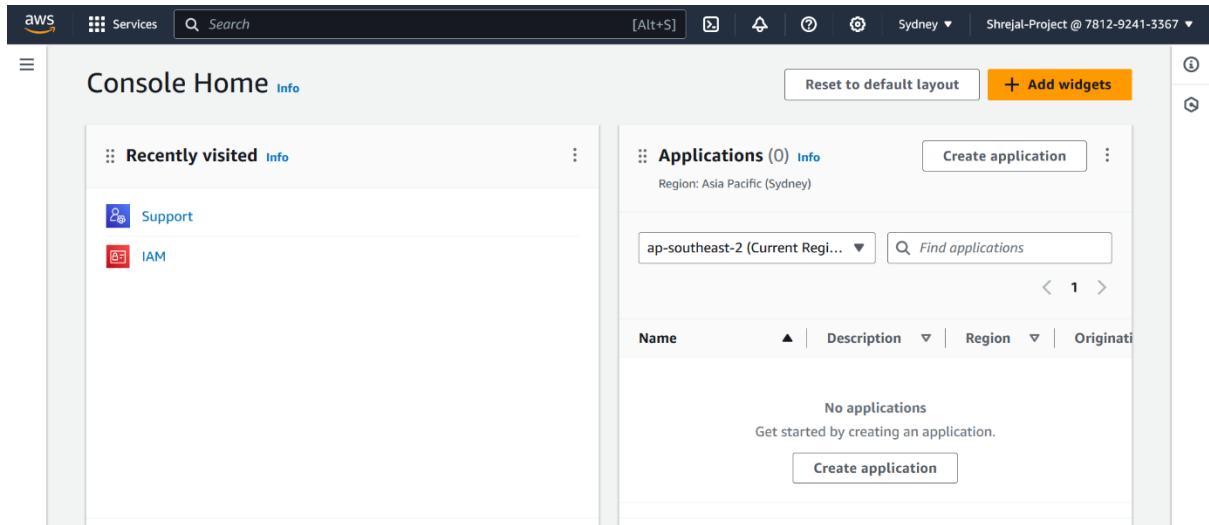
- Multiple ways to interact with our AWS environment

✓ AWS Console
✓ Programmatically
  o E.g. – AWS CLI
  o SDK

- AWS Region

  ✓ Region - physical location around the world where AWS has data centers
  ✓ AZ - one or more discrete data centers with redundant power networking and connectivity in an AWS region.

- Step-3: Install AWS CLI

  HTTPS://AWS.AMAZON.COM/CLI/