

REAL-TIME NETFLIX DATA STREAMING ON AWS USING APACHE KAFKA, ZOOKEEPER AND FLASK

PRESENTATION BY

EMEKA OKEKE

PROJECT SUMMARY

SET UP AWS
ENVIRONMENT

SYSTEM ARCHITECTURE

DATA ACQUISITION

IMPLEMENT KAFKA
CONSUMERS

INSTALL KAFKA &
ZOOKEEPER

DEVELOP KAFKA
PRODUCER

FLASK WEB INTERFACE

- ❑ Implementing a real-time data streaming solution for Netflix data on Amazon Web Services (AWS) through the utilization of Apache Kafka, Zookeeper and Flask interface presents an exciting and complex endeavor for group diving into the world of big data and cloud computing. Apache Kafka serves as the backbone of this streaming architecture, acting as a distributed streaming platform capable of handling vast amounts of data in real-time.
- ❑ The project aims to leverage cutting-edge technologies, enhancing both technical proficiency and the ability to address the growing demand for real-time analytics in the streaming industry. By leveraging Flask for the user interface, this project will not only ensure technical proficiency but also market differentiation and strategic growth opportunities.



[←](#)
[↺](#)
[🔒](#)
<https://www.kaggle.com/datasets/shivamb/netflix-shows>

[🏠](#)
[A](#)
[☆](#)
[📄](#)
[🌟](#)
[🔒](#)
[🌱](#)
[👤](#)
[⋮](#)

[+](#) Create

[🏠 Home](#)
[🏆 Competitions](#)
[📁 Datasets](#)
[🤖 Models](#)
[🔗 Code](#)
[💬 Discussions](#)
[🎓 Learn](#)
[⌵ More](#)

Search

SHIVAM BANSAL · UPDATED 2 YEARS AGO

[▲ 8020](#)
[New Notebook](#)
[Download \(1 MB\)](#)

Netflix Movies and TV Shows

Listings of movies and tv shows on Netflix - Regularly Updated

[Data Card](#)
[Code \(1465\)](#)
[Discussion \(76\)](#)

About Dataset

Other Platform's Datasets (Click on the logos to view)

- [Amazon Prime Video Movies and TV Shows](#)
- [Disney+ Movies and TV Shows](#)
- [Netflix Prime Video Movies and TV Shows](#)
- [Hulu Movies and TV Shows](#)

Netflix Movies and TV Shows

About this Dataset: Netflix is one of the most popular media and video streaming platforms. They have over 8000 movies or tv shows available on their platform as of mid-2021. they have over 200M Subscribers globally. This tabular dataset consists of listings of all the

Usability ⓘ
10.00

License
[CC0: Public Domain](#)

Expected update frequency
Quarterly

Tags

[Arts and Entertainment](#)
[Movies and TV Shows](#)
[Data Visualization](#)
[Exploratory Data Analysis](#)
[Websites](#)



DATA AQUISITION

A1																									
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W		
1	show_id	type	title	director	cast	country	date_added	release_ye	rating	duration	listed_in	description													
2	s1	Movie	Dick Johns	Kirsten Johnson		United Sta	25-Sep-21	2020	PG-13	90 min	Document	As her father nears the end of his life, filmmaker Kirsten Johnson stages his death in inventive and comical ways to help them both f													
3	s2	TV Show	Blood & Water		Ama Qama	South Afric	24-Sep-21	2021	TV-MA	2 Seasons	Internatio	After crossing paths at a party, a Cape Town teen sets out to prove whether a private-school swimming star is her sister who was a													
4	s3	TV Show	Ganglands	Julien Lecl	Sami Bouajila, Tracy C		24-Sep-21	2021	TV-MA	1 Season	Crime TV S	To protect his family from a powerful drug lord, skilled thief Mehdi and his expert team of robbers are pulled into a violent and dea													
5	s4	TV Show	Jailbirds	New Orleans			24-Sep-21	2021	TV-MA	1 Season	Docuserie	Feuds, flirtations and toilet talk go down among the incarcerated women at the Orleans Justice Center in New Orleans on this gritty													
6	s5	TV Show	Kota Factory		Mayur Mo	India	24-Sep-21	2021	TV-MA	2 Seasons	Internatio	In a city of coaching centers known to train Indiaâ€™s finest collegiate minds, an earnest but unexceptional student and his friends													
7	s6	TV Show	Midnight M	Mike Flana	Kate Siegel, Zach Gilfo		24-Sep-21	2021	TV-MA	1 Season	TV Dramas	The arrival of a charismatic young priest brings glorious miracles, ominous mysteries and renewed religious fervor to a dying town d													
8	s7	Movie	My Little P	Robert Cul	Vanessa Hudgens, Kin		24-Sep-21	2021	PG	91 min	Children &	Equestria's divided. But a bright-eyed hero believes Earth Ponies, Pegasi and Unicorns should be pals â€™ and, hoof to heart, sheâ€™													
9	s8	Movie	Sankofa	Haile Gerir	Kofi Ghanâ	United Sta	24-Sep-21	1993	TV-MA	125 min	Dramas, In	On a photo shoot in Ghana, an American model slips back in time, becomes enslaved on a plantation and bears witness to the agon													
10	s9	TV Show	The Great	Andy Devco	Mel Giedre	United Kin	24-Sep-21	2021	TV-14	9 Seasons	British TV	A talented batch of amateur bakers face off in a 10-week competition, whipping up their best dishes in the hopes of being named th													
11	s10	Movie	The Starlin	Theodore	Melissa M	United Sta	24-Sep-21	2021	PG-13	104 min	Comedies,	A woman adjusting to life after a loss contends with a feisty bird that's taken over her garden â€™ and a husband who's struggling to													
12	s11	TV Show	Vendetta: Truth	Lies and The Mafia			24-Sep-21	2021	TV-MA	1 Season	Crime TV S	Sicily boasts a bold "Anti-Mafia" coalition. But what happens when those trying to bring down organized crime are accused of being													
13	s12	TV Show	Bangkok B	Kongkiat K	Sukollawat	Kanarot, S	23-Sep-21	2021	TV-MA	1 Season	Crime TV S	Struggling to earn a living in Bangkok, a man joins an emergency rescue service and realizes he must unravel a citywide conspiracy.													
14	s13	Movie	Je Suis Kar	Christian S	Luna Wedl	Germany, I	23-Sep-21	2021	TV-MA	127 min	Dramas, In	After most of her family is murdered in a terrorist bombing, a young woman is unknowingly lured into joining the very group that kill													
15	s14	Movie	Confessor	Bruno Gar	Klara Castanho, Lucca		22-Sep-21	2021	TV-PG	91 min	Children &	When the clever but socially-awkward Tetã joins a new school, she'll do anything to fit in. But the queen bee among her classmate													
16	s15	TV Show	Crime Stories: India	Detectives			22-Sep-21	2021	TV-MA	1 Season	British TV	Cameras following Bengaluru police on the job offer a rare glimpse into the complex and challenging inner workings of four major c													
17	s16	TV Show	Dear White People		Logan Bro	United Sta	22-Sep-21	2021	TV-MA	4 Seasons	TV Comed	Students of color navigate the daily slights and slippery politics of life at an Ivy League college that's not nearly as "post-racial" as it													
18	s17	Movie	Europe's N	Pedro de Echave Garcã	ã, Pablo		22-Sep-21	2020	TV-MA	67 min	Document	Declassified documents reveal the post-WWII life of Otto Skorzeny, a close Hitler ally who escaped to Spain and became an adviser													
19	s18	TV Show	Falsa identidad		Luis Ernest	Mexico	22-Sep-21	2020	TV-MA	2 Seasons	Crime TV S	Strangers Diego and Isabel flee their home in Mexico and pretend to be a married couple to escape his drug-dealing enemies and he													
20	s19	Movie	Intrusion	Adam Salk	Freida Pinto, Logan M		22-Sep-21	2021	TV-14	94 min	Thrillers	After a deadly home invasion at a coupleâ€™s new dream house, the traumatized wife searches for answers â€™ and learns the rea													
21	s20	TV Show	Jaguar		Blanca Suã	ãjrez, Ivã	ãjn	22-Sep-21	2021	TV-MA	1 Season	Internatio	In the 1960s, a Holocaust survivor joins a group of self-trained spies who seek justice against Nazis fleeing to Spain to hide after WW												
22	s21	TV Show	Monsters I	Olivier Megaton			22-Sep-21	2021	TV-14	1 Season	Crime TV S	In the late 1970s, an accused serial rapist claims multiple personalities control his behavior, setting off a legal odyssey that captivat													
23	s22	TV Show	Resurrection: Ertugru	Engin Altar	Turkey		22-Sep-21	2018	TV-14	5 Seasons	Internatio	When a good deed unwittingly endangers his clan, a 13th-century Turkish warrior agrees to fight a sultan's enemies in exchange for													
24	s23	Movie	Avvai Shan	K.S. Raviku	Kamal Hassan, Meena		21-Sep-21	1996	TV-PG	161 min	Comedies,	Newly divorced and denied visitation rights with his daughter, a doting father disguises himself as a gray-haired nanny in order to sp													
25	s24	Movie	Go! Go! C	Alex Woo,	Maisie Benson, Paul K		21-Sep-21	2021	TV-Y	61 min	Children &	From arcade games to sled days and hiccup cures, Cory Carsonâ€™s curious little sister Chrissy speeds off on her own for fun and a													
26	s25	Movie	Jeans		S. Shankar	Prashanth, India	21-Sep-21	1998	TV-14	166 min	Comedies,	When the father of the man she loves insists that his twin sons marry twin sisters, a woman creates an alter ego that might be a bit													
27	s26	TV Show	Love on the Spectrum		Brooke Sa	Australia	21-Sep-21	2021	TV-14	2 Seasons	Docuserie	Finding love can be hard for anyone. For young adults on the autism spectrum, exploring the unpredictable world of dating is even r													
28	s27	Movie	Minsara Ka	Rajiv Meno	Arvind Swamy, Kajol,		21-Sep-21	1997	TV-PG	147 min	Comedies,	A tangled love triangle ensues when a man falls for a woman studying to become a nun â€™ and she falls for the friend he enlists to													
29	s28	Movie	Grown Up:	Dennis Du	Adam Sanc	United Sta	20-Sep-21	2010	PG-13	103 min	Comedies	Mourning the loss of their beloved junior high basketball coach, five middle-aged pals reunite at a lake house and rediscover the joy													
30	s29	Movie	Dark Skies	Scott Stew	Keri Russel	United Sta	19-Sep-21	2013	PG-13	97 min	Horror Mo	A familyâ													
31	s30	Movie	Paranoia	Robert Luk	Liam Hem	United Sta	19-Sep-21	2013	PG-13	106 min	Thrillers	Blackmailed by his company's CEO, a low-level employee finds himself forced to spy on the boss's rival and former mentor.													
32	s31	Movie	Ankahi Ka	Ashwiny Iy	Abhishek Banerjee, Ri		17-Sep-21	2021	TV-14	111 min	Dramas, In	As big city life buzzes around them, lonely souls discover surprising sources of connection and companionship in three tales of love,													
33	s32	TV Show	Chicago Party Aunt		Lauren Ash, Rory O'M		17-Sep-21	2021	TV-MA	1 Season	TV Comed	Chicago Party Aunt Diane is an idolized troublemaker with a talent for avoiding adulthood â€™ and a soft spot for her soul-searching													

RETURN

Search results for 'EC2'

Try searching with longer queries for more relevant results

Services (13)

Features (54)


Resources **New**

Documentation (34,127)

Knowledge Articles (598)

Marketplace (7,370)

Services [See all 13 results ►](#)

 **EC2** ☆
Virtual Servers in the Cloud

[EC2](#) > [Instances](#) > Launch an instance

Launch an instance [Info](#)

Amazon EC2 allows you to create virtual machines, or instances, that run on the AWS Cloud. Quickly get started by following the simple steps below.

Name and tags [Info](#)



Name



[Add additional tags](#)



▼ Application and OS Images (Amazon Machine Image) [Info](#)



An AMI is a template that contains the software configuration (operating system, application server, and applications) required to launch your instance. Search or Browse for AMIs if you don't see what you are looking for below



Quick Start




Amazon Linux




Ubuntu



Windows



Red Hat



SUSE Linux



Debian



[Browse more AMIs](#)
Including AMIs from AWS, Marketplace and the Community

Amazon Machine Image (AMI)

▼ Summary

Number of instances [Info](#)

Software Image (AMI)

Amazon Linux 2023 AMI 2023.2.2...[read more](#)
ami-095819c19b51bc983

Virtual server type (instance type)



t2.micro

Firewall (security group)

New security group

Storage (volumes)

1 volume(s) - 8 GiB

 **Free tier:** In your first year includes 750 hours of t2.micro (or t3.micro in the Regions in which t2.micro is unavailable) instance usage on free tier AMIs per month, 30 GiB of EBS storage, 2 million IOs, 1 GB of snapshots, and 100 GB of bandwidth to the internet. 

[Cancel](#)[Launch instance](#)[Review commands](#)

RETURN

Create key pair

Key pair name

Key pairs allow you to connect to your instance securely.

netflix-daty

The name can include upto 255 ASCII characters. It can't include leading or trailing spaces.

Key pair type


☒ RSA
RSA encrypted private and public key pair

☐ ED25519
ED25519 encrypted private and public key pair

Private key file format

☒ .pem
For use with OpenSSH

☐ .ppk
For use with PuTTY

 When prompted, store the private key in a secure and accessible location on your computer. **You will need it later to connect to your instance.** [Learn more](#)

Cancel

Create key pair



SET UP AWS ENVIRONMENT

AWS Services Search [Alt+S]

EC2 Dashboard X
EC2 Global View
Events

▼ Instances
Instances
Instance Types
Launch Templates
Spot Requests
Savings Plans
Reserved Instances
Dedicated Hosts
Capacity Reservations
[New](#)

▼ Images
AMIs
AMI Catalog

▼ Elastic Block Store
Volumes
Snapshots
Lifecycle Manager

▼ Network & Security
Security Groups
Elastic IPs
Placement Groups
Key Pairs
Network Interfaces

Successfully started i-058f0855c8addea28

Instances (1/1) Info

Find Instance by attribute or tag (case-sensitive)

<input checked="" type="checkbox"/>	Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IPv4 DNS	Public IPv4 ...	Elastic IP
<input checked="" type="checkbox"/>	netflix-data	i-058f0855c8addea28	Pending	t2.micro	-	No alarms +	ca-central-1a	-	-	-

Instance: i-058f0855c8addea28 (netflix-data)

Details | Security | Networking | Storage | Status checks | Monitoring | Tags

▼ Instance summary Info

Instance ID
i-058f0855c8addea28 (netflix-data)

IPv6 address
-

Hostname type
IP name: ip-172-31-30-187.ca-central-1.compute.internal

Answer private resource DNS name
IPv4 (A)
Auto-assigned IP address

Public IPv4 address
-

Instance state
Pending

Private IP DNS name (IPv4 only)
ip-172-31-30-187.ca-central-1.compute.internal

Instance type
t2.micro

VPC ID

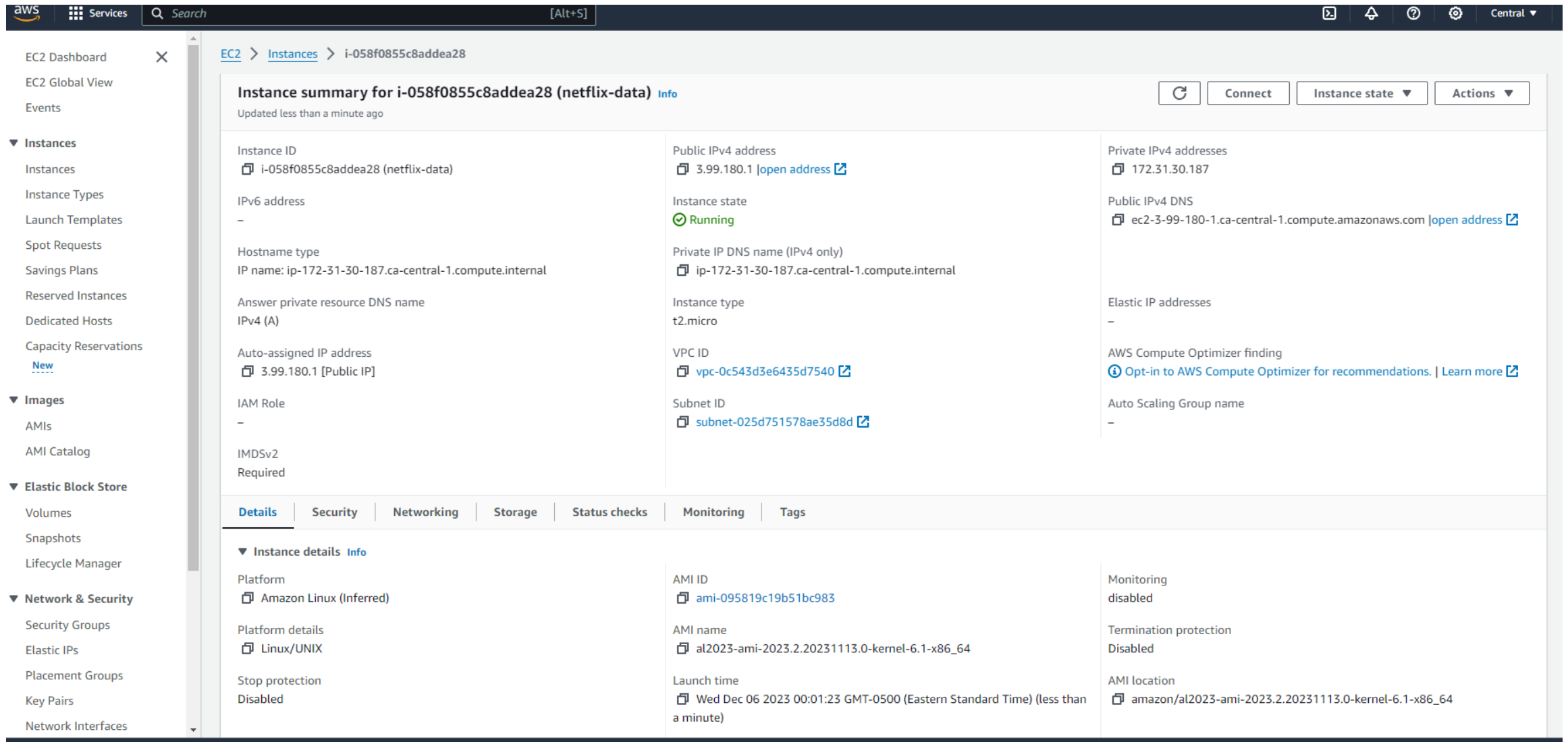
Private IPv4 addresses
172.31.30.187

Public IPv4 DNS
-

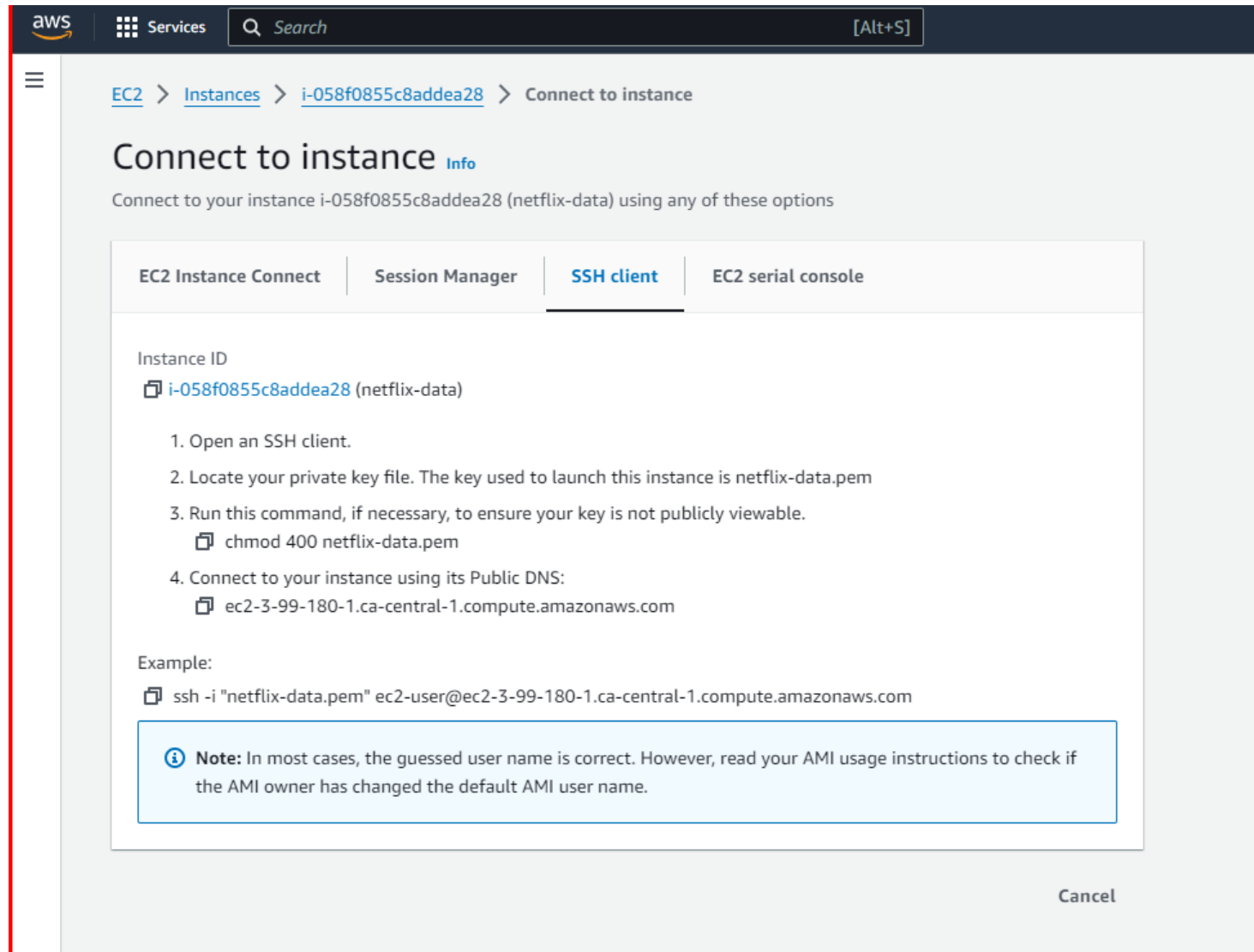
Elastic IP addresses
-

AWS Compute Optimizer finding

RETURN



RETURN



The screenshot shows the AWS Management Console interface. At the top, there's a navigation bar with the AWS logo, 'Services' menu, a search bar, and a keyboard shortcut '[Alt+S]'. Below this, a breadcrumb trail reads 'EC2 > Instances > i-058f0855c8addea28 > Connect to instance'. The main heading is 'Connect to instance' with an 'Info' link. A subtext says 'Connect to your instance i-058f0855c8addea28 (netflix-data) using any of these options'. There are four tabs: 'EC2 Instance Connect', 'Session Manager', 'SSH client' (which is selected), and 'EC2 serial console'. Under the 'SSH client' tab, the 'Instance ID' is listed as 'i-058f0855c8addea28 (netflix-data)'. A list of four steps is provided: 1. Open an SSH client. 2. Locate your private key file. The key used to launch this instance is netflix-data.pem. 3. Run this command, if necessary, to ensure your key is not publicly viewable. A code block shows 'chmod 400 netflix-data.pem'. 4. Connect to your instance using its Public DNS: A code block shows 'ec2-3-99-180-1.ca-central-1.compute.amazonaws.com'. Below the steps, an 'Example:' section shows a terminal command: 'ssh -i "netflix-data.pem" ec2-user@ec2-3-99-180-1.ca-central-1.compute.amazonaws.com'. A blue-bordered box contains a note: 'Note: In most cases, the guessed user name is correct. However, read your AMI usage instructions to check if the AMI owner has changed the default AMI user name.' At the bottom right, there is a 'Cancel' button.

aws Services Search [Alt+S]

EC2 > Instances > i-058f0855c8addea28 > Connect to instance

Connect to instance [Info](#)

Connect to your instance i-058f0855c8addea28 (netflix-data) using any of these options


EC2 Instance Connect



Session Manager

SSH client


EC2 serial console


Instance ID

 i-058f0855c8addea28 (netflix-data)

1. Open an SSH client.
2. Locate your private key file. The key used to launch this instance is netflix-data.pem
3. Run this command, if necessary, to ensure your key is not publicly viewable.
 `chmod 400 netflix-data.pem`
4. Connect to your instance using its Public DNS:
 `ec2-3-99-180-1.ca-central-1.compute.amazonaws.com`

Example:

 `ssh -i "netflix-data.pem" ec2-user@ec2-3-99-180-1.ca-central-1.compute.amazonaws.com`

 **Note:** In most cases, the guessed user name is correct. However, read your AMI usage instructions to check if the AMI owner has changed the default AMI user name.

Cancel

netflix-data.pem

12/4/2023 5:58 PM

PEM File

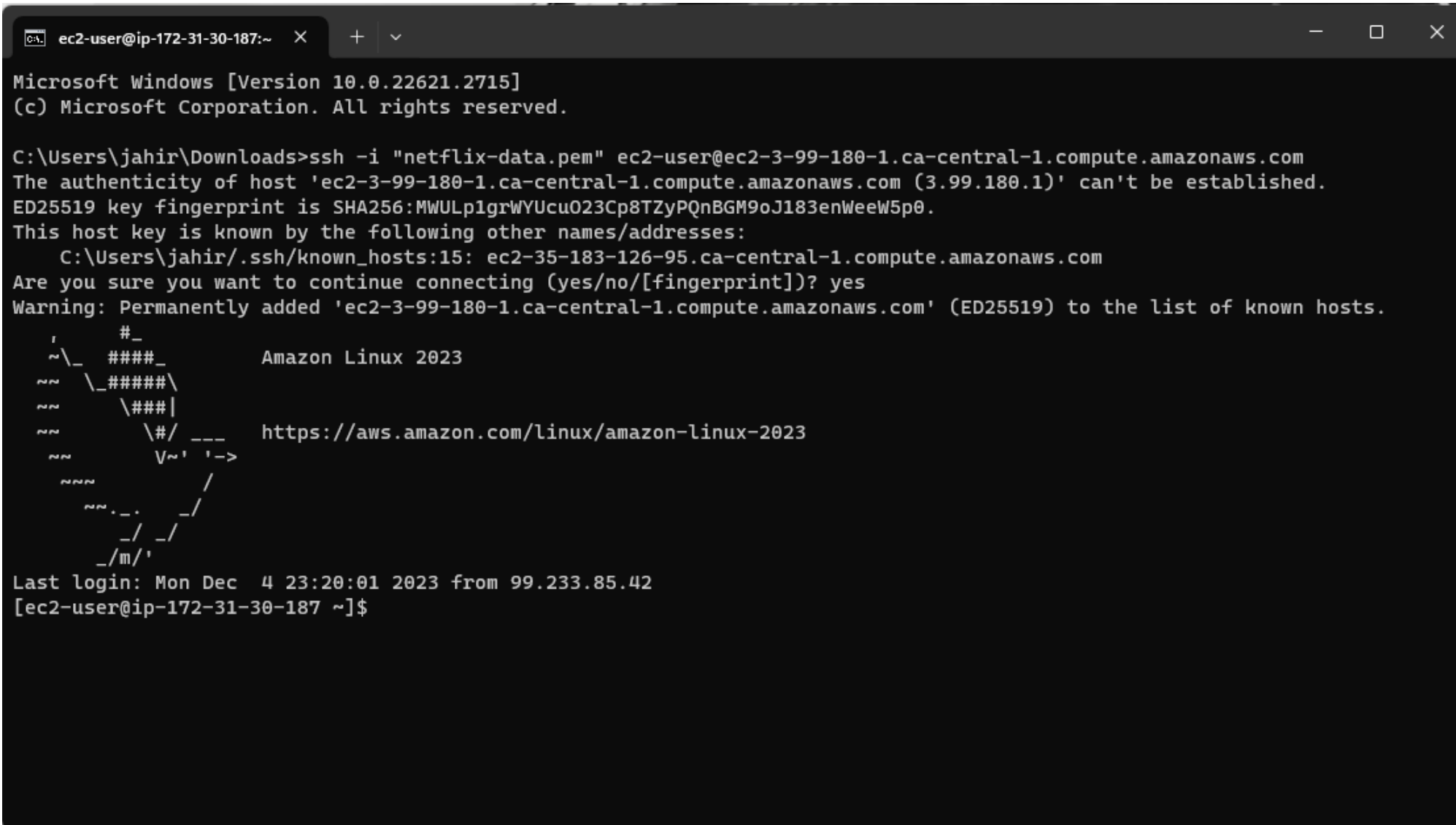
2 KB

```
C:\Windows\System32\cmd.e X + v
Microsoft Windows [Version 10.0.22621.2715]
(c) Microsoft Corporation. All rights reserved.

C:\Users\jahir\Downloads>ssh -i "netflix-data.pem" ec2-user@ec2-3-99-180-1.ca-central-1.compute.amazonaws.com
```

```
C:\Windows\System32\cmd.e X + v
Microsoft Windows [Version 10.0.22621.2715]
(c) Microsoft Corporation. All rights reserved.

C:\Users\jahir\Downloads>ssh -i "netflix-data.pem" ec2-user@ec2-3-99-180-1.ca-central-1.compute.amazonaws.com
The authenticity of host 'ec2-3-99-180-1.ca-central-1.compute.amazonaws.com (3.99.180.1)' can't be established.
ED25519 key fingerprint is SHA256:MWULp1grWYUcu023Cp8TZyPQnBGM9oJl83enWeeW5p0.
This host key is known by the following other names/addresses:
  C:\Users\jahir/.ssh/known_hosts:15: ec2-35-183-126-95.ca-central-1.compute.amazonaws.com
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
```



RETURN

https://georgiancollege-my.sharepoint.com/:t:/g/personal/200546259_student_georgi_anc_on_ca/Ed8Hr_9KxBxGtJRBilkey_sB-ZQBPX405EyxWRX3v4KbPA?e=yqRhV

https://georgiancollege-my.sharepoint.com/:t:/g/personal/200546259_student_georgi_anc_on_ca/EXbCDWaVpaZHn6xPy7RwKEABo9FTusWHKuMAgLyfezSC8g?e=kAueFq

```

[ec2-user@ip-172-31-30-187:~] x + v
The authenticity of host 'ec2-3-99-180-1.ca-central-1.compute.amazonaws.com (3.99.180.1)' can't be established.
ED25519 key fingerprint is SHA256:MWULp1grWYUcu023Cp8TZyPQnBGM9oJ183enWeeW5p0.
This host key is known by the following other names/addresses:
  C:\Users\jahir/.ssh/known_hosts:15: ec2-35-183-126-95.ca-central-1.compute.amazonaws.com
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added 'ec2-3-99-180-1.ca-central-1.compute.amazonaws.com' (ED25519) to the list of known hosts.

      #_
    ~\_  ####_      Amazon Linux 2023
  ~ ~ \_#####\
  ~ ~  \###|
  ~ ~   \#/  ___  https://aws.amazon.com/linux/amazon-linux-2023
  ~ ~    V~'  '->
  ~ ~
  ~ ~ _..
  ~ ~ _/_/_/
  ~ ~ _/m/'

Last login: Mon Dec  4 23:20:01 2023 from 99.233.85.42
[ec2-user@ip-172-31-30-187 ~]$ wget https://downloads.apache.org/kafka/3.6.0/kafka_2.12-3.6.0.tgz
--2023-12-06 05:09:51-- https://downloads.apache.org/kafka/3.6.0/kafka_2.12-3.6.0.tgz
Resolving downloads.apache.org (downloads.apache.org)... 88.99.95.219, 135.181.214.104, 2a01:4f9:3a:2c57::2, ...
Connecting to downloads.apache.org (downloads.apache.org)|88.99.95.219|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 113400977 (108M) [application/x-gzip]
Saving to: 'kafka_2.12-3.6.0.tgz.1'

kafka_2.12-3.6.0.tgz.1      100%[=====>] 108.15M  14.6MB/s   in 8.4s

2023-12-06 05:10:00 (12.9 MB/s) - 'kafka_2.12-3.6.0.tgz.1' saved [113400977/113400977]

[ec2-user@ip-172-31-30-187 ~]$

```



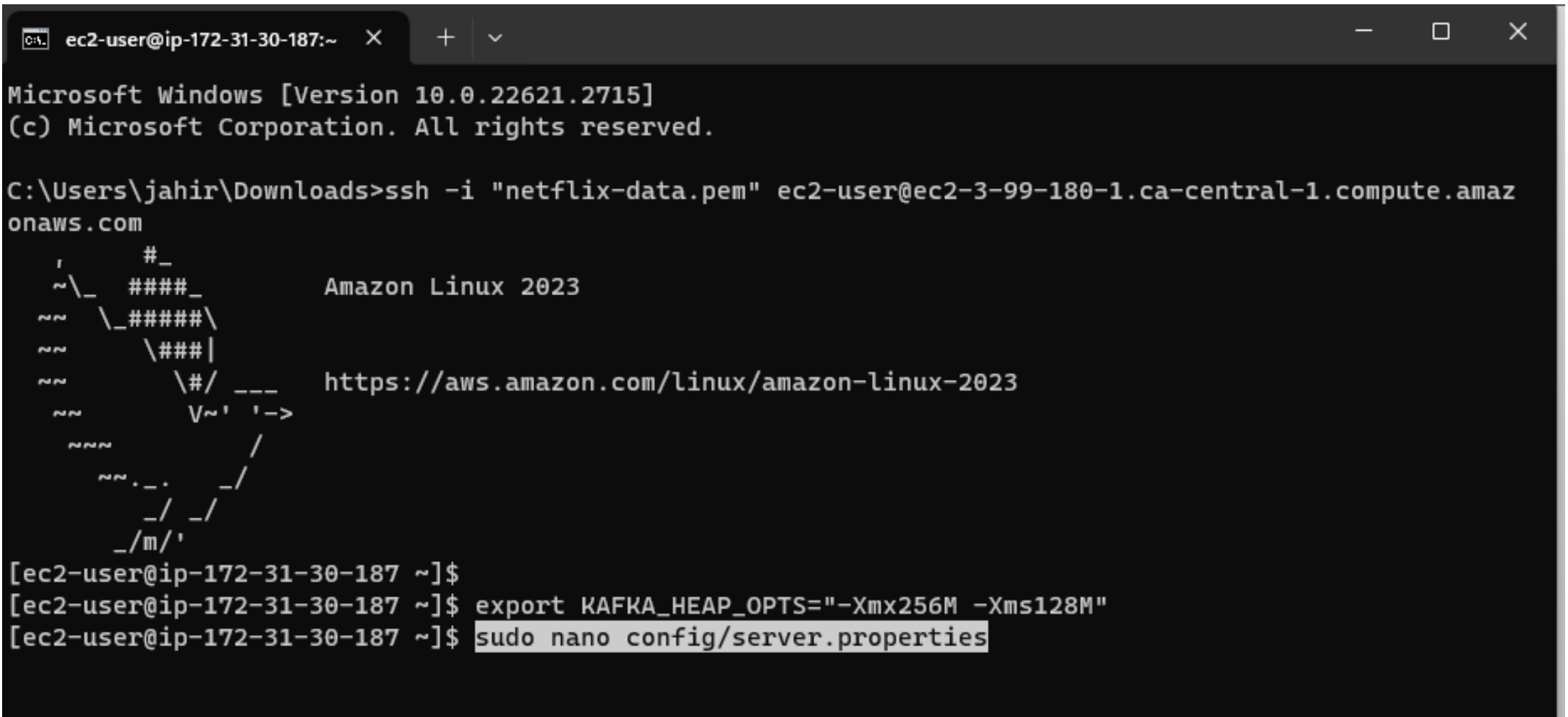
```
[ec2-user@ip-172-31-30-187 ~]$ java -version
openjdk version "17.0.9" 2023-10-17 LTS
OpenJDK Runtime Environment Corretto-17.0.9.8.1 (build 17.0.9+8-LTS)
OpenJDK 64-Bit Server VM Corretto-17.0.9.8.1 (build 17.0.9+8-LTS, mixed mode, sharing)
[ec2-user@ip-172-31-30-187 ~]$
```

```
[ec2-user@ip-172-31-30-187 ~]$ java -version
openjdk version "17.0.9" 2023-10-17 LTS
OpenJDK Runtime Environment Corretto-17.0.9.8.1 (build 17.0.9+8-LTS)
OpenJDK 64-Bit Server VM Corretto-17.0.9.8.1 (build 17.0.9+8-LTS, mixed mode, sharing)
[ec2-user@ip-172-31-30-187 ~]$ cd kafka_2.12-3.6.0
[ec2-user@ip-172-31-30-187 kafka_2.12-3.6.0]$
```

```

2023-12-06 05:13:06,512] INFO metricsProvider.className is org.apache.zookeeper.metrics.impl.DefaultMetricsProvider (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
2023-12-06 05:13:06,517] INFO autopurge.snapRetainCount set to 3 (org.apache.zookeeper.server.DatadirCleanupManager)
2023-12-06 05:13:06,517] INFO autopurge.purgeInterval set to 0 (org.apache.zookeeper.server.DatadirCleanupManager)
2023-12-06 05:13:06,517] INFO Purge task is not scheduled. (org.apache.zookeeper.server.DatadirCleanupManager)
2023-12-06 05:13:06,517] WARN Either no config or no quorum defined in config, running in standalone mode (org.apache.zookeeper.server.quorum.QuorumPeerMain)
2023-12-06 05:13:06,521] INFO Log4j 1.2 jmx support not found; jmx disabled. (org.apache.zookeeper.jmx.ManagedUtil)
2023-12-06 05:13:06,523] INFO Reading configuration from: config/zookeeper.properties (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
2023-12-06 05:13:06,523] WARN config/zookeeper.properties is relative. Prepend ./ to indicate that you're sure! (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
2023-12-06 05:13:06,523] INFO clientPortAddress is 0.0.0.0:2181 (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
2023-12-06 05:13:06,524] INFO secureClientPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
2023-12-06 05:13:06,524] INFO observerMasterPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
2023-12-06 05:13:06,524] INFO metricsProvider.className is org.apache.zookeeper.metrics.impl.DefaultMetricsProvider (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
2023-12-06 05:13:06,524] INFO Starting server (org.apache.zookeeper.server.ZooKeeperServerMain)
2023-12-06 05:13:06,561] INFO ServerMetrics initialized with provider org.apache.zookeeper.metrics.impl.DefaultMetricsProvider@161b062a (org.apache.zookeeper.server.ServerMetrics)
2023-12-06 05:13:06,578] INFO ACL digest algorithm is: SHA1 (org.apache.zookeeper.server.auth.DigestAuthenticationProvider)
2023-12-06 05:13:06,579] INFO zookeeper.DigestAuthenticationProvider.enabled = true (org.apache.zookeeper.server.auth.DigestAuthenticationProvider)
2023-12-06 05:13:06,589] INFO zookeeper.snapshot.trust.empty : false (org.apache.zookeeper.server.persistence.FileTxnSnapLog)
2023-12-06 05:13:06,617] INFO (org.apache.zookeeper.server.ZooKeeperServer)
2023-12-06 05:13:06,617] INFO (org.apache.zookeeper.server.ZooKeeperServer)
2023-12-06 05:13:06,617] INFO (org.apache.zookeeper.server.ZooKeeperServer)
2023-12-06 05:13:06,618] INFO (org.apache.zookeeper.server.ZooKeeperServer)
2023-12-06 05:13:06,618] INFO (org.apache.zookeeper.server.ZooKeeperServer)
2023-12-06 05:13:06,618] INFO (org.apache.zookeeper.server.ZooKeeperServer)
2023-12-06 05:13:06,618] INFO (org.apache.zookeeper.server.ZooKeeperServer)
2023-12-06 05:13:06,618] INFO (org.apache.zookeeper.server.ZooKeeperServer)
2023-12-06 05:13:06,618] INFO (org.apache.zookeeper.server.ZooKeeperServer)
2023-12-06 05:13:06,619] INFO (org.apache.zookeeper.server.ZooKeeperServer)
2023-12-06 05:13:06,619] INFO (org.apache.zookeeper.server.ZooKeeperServer)
2023-12-06 05:13:06,624] INFO Server environment:zookeeper.version=3.8.2-139d619b58292d7734b4fc83a0f44be4e7b0c986, built on 2023-07-05 19:24 UTC (org.apache.zookeeper.server.ZooKeeperServer)
2023-12-06 05:13:06,625] INFO Server environment:host.name=ip-172-31-30-187.ca-central-1.compute.internal (org.apache.zookeeper.server.ZooKeeperServer)
2023-12-06 05:13:06,625] INFO Server environment:java.version=17.0.9 (org.apache.zookeeper.server.ZooKeeperServer)
2023-12-06 05:13:06,625] INFO Server environment:java.vendor=Amazon.com Inc. (org.apache.zookeeper.server.ZooKeeperServer)
2023-12-06 05:13:06,625] INFO Server environment:java.home=/usr/lib/jvm/java-17-amazon-corretto.x86_64 (org.apache.zookeeper.server.ZooKeeperServer)
2023-12-06 05:13:06,625] INFO Server environment:java.class.path=/home/ec2-user/kafka_2.12-3.6.0/bin/../libs/activation-1.1.1.jar:/home/ec2-user/kafka_2.12-3.6.0/bin/../lib
home/ec2-user/kafka_2.12-3.6.0/bin/../libs/argparse4j-0.7.0.jar:/home/ec2-user/kafka_2.12-3.6.0/bin/../libs/audience-annotations-0.12.0.jar:/home/ec2-user/kafka_2.12-3.6.0/b

```



RETURN



INSTALL KAFKA & ZOOKEEPER



aws

Services

Search

[Alt+S]

EC2 Dashboard

EC2 Global View

Events

Instances

Instances

Instance Types

Launch Templates

Spot Requests

Savings Plans

Reserved Instances

Dedicated Hosts

Capacity Reservations

Images

AMIs

AMI Catalog

Elastic Block Store

Volumes

Snapshots

Lifecycle Manager

Network & Security

Security Groups

Elastic IPs

Placement Groups

Key Pairs

Network Interfaces

EC2 > Instances > i-058f0855c8addea28

Instance summary for i-058f0855c8addea28 (netflix-data) Info

Updated less than a minute ago

Instance ID

i-058f0855c8addea28 (netflix-data)

IPv6 address

-

Hostname type

IP name: ip-172-31-30-187.ca-central-1.compute.internal

Answer private resource DNS name

IPv4 (A)

Auto-assigned IP address

3.99.180.1 [Public IP]

IAM Role

-

IMDSv2

Required

Public IPv4 address

3.99.180.1 |open address

Instance state

Running

Private IP DNS name (IPv4 only)

Private IPv4 addresses

172.31.30.187

Public IPv4 DNS

ec2-3-99-180-1.ca-central-1.compute.amazonaws.com |open address

Details

Security

Networking

Storage

Status checks

▼ Instance details Info

Platform

Amazon Linux (Inferred)

Platform details

Linux/UNIX

Stop protection

Disabled

AMI name

al2023-ami-2023.2.20231113.0-kernel-6.1-x86_64

Launch time

Wed Dec 06 2023 00:01:23 GMT-0500 (Eastern Standard Time) (16 minutes)

Termination protection

Disabled

AMI location

amazon/al2023-ami-2023.2.20231113.0-kernel-6.1-x86_64

ec2-user@ip-172-31-30-187:~\$

GNU nano 5.8

config/server.properties

Modified

broker.id=0

Socket Server Settings

The address the socket server listens on. If not configured, the host name will be equal to the v

java.net.InetAddress.getCanonicalHostName(), with PLAINTEXT listener name, and port 9092.

#

FORMAT:

listeners = listener_name://host_name:port

#

EXAMPLE:

listeners = PLAINTEXT://your.host.name:9092

#listeners=PLAINTEXT://:9092

Listener name, hostname and port the broker will advertise to clients.

If not set, it uses the value for "listeners".

advertised.listeners=PLAINTEXT://3.99.180.1:9092

^G Help

^O Write Out

^W Where Is

^K Cut

^T Execute

^C Location

M-U Undo

^X Exit

^R Read File

^_ Replace

^U Paste

^J Justify

^_ Go To Line

M-E Redo

RETURN

```
ec2-user@ip-172-31-30-187:~$ ssh -i "netflix-data.pem" ec2-user@ec2-3-99-180-1.ca-central-1.compute.amazonaws.com
(c) Microsoft Corporation. All rights reserved.

C:\Users\jahir\Downloads>ssh -i "netflix-data.pem" ec2-user@ec2-3-99-180-1.ca-central-1.compute.amazonaws.com

#_
~\_ #####_      Amazon Linux 2023
~~ \_#####\
~~  \###|
~~   \#/ _-- https://aws.amazon.com/linux/amazon-linux-2023
~~    V~' '->
~~~
~~._. _/ _/
   _/_/_/
   _/m/'

Last login: Wed Dec  6 05:04:38 2023 from 99.233.85.42
[ec2-user@ip-172-31-30-187 ~]$ export KAFKA_HEAP_OPTS="-Xmx256M -Xms128M"
[ec2-user@ip-172-31-30-187 ~]$ sudo nano config/server.properties
[ec2-user@ip-172-31-30-187 ~]$ cd kafka_2.12-3.6.0
[ec2-user@ip-172-31-30-187 kafka_2.12-3.6.0]$ sudo nano config/server.properties
[ec2-user@ip-172-31-30-187 kafka_2.12-3.6.0]$ bin/kafka-server-start.sh config/server.properties
```



INSTALL KAFKA & ZOOKEEPER



aws

Services

Search

[Alt+S]

EC2 Dashboard

EC2 Global View

Events

Instances

Instances

Instance Types

Launch Templates

Spot Requests

Savings Plans

Reserved Instances

Dedicated Hosts

Capacity Reservations

Images

AMIs

AMI Catalog

Elastic Block Store

Volumes

Snapshots

Lifecycle Manager

Network & Security

Security Groups

Elastic IPs

Placement Groups

Key Pairs

Network Interfaces

EC2 > Instances > i-058f0855c8addea28

ec2-user@ip-172-31-30-187:~

+

▼

```
[2023-12-06 05:21:37,937] INFO [ExpirationReaper-0-Produce]: Starting (kafka.server.DelayedOperationPurgatory$ExpiredOperationReaper)
[2023-12-06 05:21:37,940] INFO [ExpirationReaper-0-Fetch]: Starting (kafka.server.DelayedOperationPurgatory$ExpiredOperationReaper)
[2023-12-06 05:21:37,941] INFO [ExpirationReaper-0-DeleteRecords]: Starting (kafka.server.DelayedOperationPurgatory$ExpiredOperationReaper)
[2023-12-06 05:21:37,942] INFO [ExpirationReaper-0-ElectLeader]: Starting (kafka.server.DelayedOperationPurgatory$ExpiredOperationReaper)
[2023-12-06 05:21:37,951] INFO [ExpirationReaper-0-RemoteFetch]: Starting (kafka.server.DelayedOperationPurgatory$ExpiredOperationReaper)
[2023-12-06 05:21:37,977] INFO [LogDirFailureHandler]: Starting (kafka.server.ReplicaManager$LogDirFailureHandler)
[2023-12-06 05:21:37,982] INFO [AddPartitionsToTxnSenderThread-0]: Starting (kafka.server.AddPartitionsToTxnManager)
[2023-12-06 05:21:38,016] INFO Creating /brokers/ids/0 (is it secure? false) (kafka.zk.KafkaZkClient)
[2023-12-06 05:21:38,061] INFO Stat of the created znode at /brokers/ids/0 is: 25,25,1701840098046,1701840098046,1,0,0,72057638974259200,204,0,25
(kafka.zk.KafkaZkClient)
[2023-12-06 05:21:38,062] INFO Registered broker 0 at path /brokers/ids/0 with addresses: PLAINTEXT://3.99.180.1:9092, czxid (broker epoch): 25 (kafka.zk.KafkaZkClient)
[2023-12-06 05:21:38,188] INFO [ExpirationReaper-0-topic]: Starting (kafka.server.DelayedOperationPurgatory$ExpiredOperationReaper)
[2023-12-06 05:21:38,201] INFO Successfully created /controller_epoch with initial epoch 0 (kafka.zk.KafkaZkClient)
[2023-12-06 05:21:38,231] INFO Feature ZK node created at path: /feature (kafka.server.FinalizedFeatureChangeListener)
[2023-12-06 05:21:38,245] INFO [ExpirationReaper-0-Heartbeat]: Starting (kafka.server.DelayedOperationPurgatory$ExpiredOperationReaper)
[2023-12-06 05:21:38,246] INFO [ExpirationReaper-0-Rebalance]: Starting (kafka.server.DelayedOperationPurgatory$ExpiredOperationReaper)
[2023-12-06 05:21:38,294] INFO [GroupCoordinator 0]: Starting up. (kafka.coordinator.group.GroupCoordinator)
[2023-12-06 05:21:38,308] INFO [MetadataCache brokerId=0] Updated cache from existing None to latest Features(version=3.6-IV2, finalizedFeatures={}, finalizedFeaturesEpoch=0). (kafka.server.metadata.ZkMetadataCache)
[2023-12-06 05:21:38,343] INFO [GroupCoordinator 0]: Startup complete. (kafka.coordinator.group.GroupCoordinator)
[2023-12-06 05:21:38,399] INFO [TransactionCoordinator id=0] Starting up. (kafka.coordinator.transaction.TransactionCoordinator)
[2023-12-06 05:21:38,423] INFO [TransactionCoordinator id=0] Startup complete. (kafka.coordinator.transaction.TransactionCoordinator)
[2023-12-06 05:21:38,428] INFO [TxnMarkerSenderThread-0]: Starting (kafka.coordinator.transaction.TransactionMarkerChannelManager)
[2023-12-06 05:21:38,522] INFO [ExpirationReaper-0-AlterAcls]: Starting (kafka.server.DelayedOperationPurgatory$ExpiredOperationReaper)
[2023-12-06 05:21:38,651] INFO [/config/changes-event-process-thread]: Starting (kafka.common.ZkNodeChangeNotificationListener$ChangeEventProcessThread)
[2023-12-06 05:21:38,660] INFO [SocketServer listenerType=ZK_BROKER, nodeId=0] Enabling request processing. (kafka.network.SocketServer)
[2023-12-06 05:21:38,663] INFO Awaiting socket connections on 0.0.0.0:9092. (kafka.network.DataPlaneAcceptor)
[2023-12-06 05:21:38,706] INFO Kafka version: 3.6.0 (org.apache.kafka.common.utils.AppInfoParser)
[2023-12-06 05:21:38,706] INFO Kafka commitId: 60e845626d8a465a (org.apache.kafka.common.utils.AppInfoParser)
[2023-12-06 05:21:38,706] INFO Kafka startTimeMs: 1701840098698 (org.apache.kafka.common.utils.AppInfoParser)
[2023-12-06 05:21:38,708] INFO [KafkaServer id=0] started (kafka.server.KafkaServer)
```

Instance state ▼

Actions ▼

Disabled

Wed Dec 06 2023 00:01:23 GMT-0500 (Eastern Standard Time) (16 minutes)

amazon/al2023-ami-2023.2.20231113.0-kernel-6.1-x86_64

RETURN



INSTALL KAFKA & ZOOKEEPER



aws

Services

Search

[Alt+S]

Central

EC2 Dashboard

EC2 Global View

Events

Instances

Instances

Instance Types

Launch Templates

Spot Requests

Savings Plans

Reserved Instances

Dedicated Hosts

Capacity Reservations

New

Images

AMIs

AMI Catalog

Elastic Block Store

Volumes

Snapshots

Lifecycle Manager

Network & Security

Security Groups

Elastic IPs

Placement Groups

Key Pairs

Network Interfaces

Instances (1/1) Info

Refresh

Connect

Instance state

Actions

Launch instances

Find Instance by attribute or tag (case-sensitive)

Instance state = running

Clear filters

<input checked="" type="checkbox"/>	Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IPv4 DNS	Public IPv4 ...	Elastic IP
<input checked="" type="checkbox"/>	netflix-data	i-058f0855c8addea28	Running	t2.micro	2/2 checks passed	No alarms	ca-central-1a	ec2-3-99-180-1.ca-cent...	3.99.180.1	-

Instance: i-058f0855c8addea28 (netflix-data)

DetailsSecurityNetworkingStorageStatus checksMonitoringTags

Security details

IAM Role

-

Owner ID

506011406156

Launch time

Wed Dec 06 2023 00:01:23 GMT-0500 (Eastern Standard Time)

Security groups

sg-0d7a7457a4b77009f (launch-wizard-1)

Inbound rules

Filter rules

Name	Security group rule ID	Port range	Protocol	Source	Security groups	Description
------	------------------------	------------	----------	--------	-----------------	-------------

RETURN



INSTALL KAFKA & ZOOKEEPER



aws

Services

Search

[Alt+S]

EC2 Dashboard

EC2 Global View

Events

Instances

Instances

Instance Types

Launch Templates

Spot Requests

Savings Plans

Reserved Instances

Dedicated Hosts

Capacity Reservations

New

Images

AMIs

AMI Catalog

Elastic Block Store

Volumes

Snapshots

Lifecycle Manager

Network & Security

Security Groups

Elastic IPs

Placement Groups

Key Pairs

Network Interfaces

EC2 > Security Groups > sg-0d7a7457a4b77009f - launch-wizard-1

sg-0d7a7457a4b77009f - launch-wizard-1

Actions

Details

Security group name

launch-wizard-1

Security group ID

sg-0d7a7457a4b77009f

Description

launch-wizard-1 created 2023-12-04T22:57:53.662Z

VPC ID

vpc-0c543d3e6435d7540

Owner

506011406156

Inbound rules count

2 Permission entries

Outbound rules count

1 Permission entry

Inbound rules

Outbound rules

Tags

Inbound rules (2)


Manage tags

Edit inbound rules

Search

	Name	Security group rule...	IP version	Type	Protocol	Port range	Source	Description
<input type="checkbox"/>	-	sgr-08afbc2e5680a3210	IPv4	SSH	TCP	22	0.0.0.0/0	-
<input type="checkbox"/>	-	sgr-0c283f0803d1bdac6	IPv4	All traffic	All	All	99.233.85.42/32	-

RETURN

 Services [Alt+S]

[EC2](#) > [Security Groups](#) > [sg-0d7a7457a4b77009f - launch-wizard-1](#) > Edit inbound rules


Edit inbound rules [Info](#)

Inbound rules control the incoming traffic that's allowed to reach the instance.

Inbound rules [Info](#)

Security group rule ID	Type Info	Protocol Info	Port range Info	Source Info	Description - optional Info	
sgr-08afbc2e5680a3210	<div>SSH</div>	TCP	22	<div>Custom</div> <div><input type="text" value="0.0.0.0/0"/></div>	<input type="text"/>	<div>Delete</div>
sgr-0c283f0803d1bdac6	<div>All traffic</div>	All	All	<div>Custom</div> <div><input type="text" value="99.233.85.42/32"/></div>	<input type="text"/>	<div>Delete</div>

Add rule

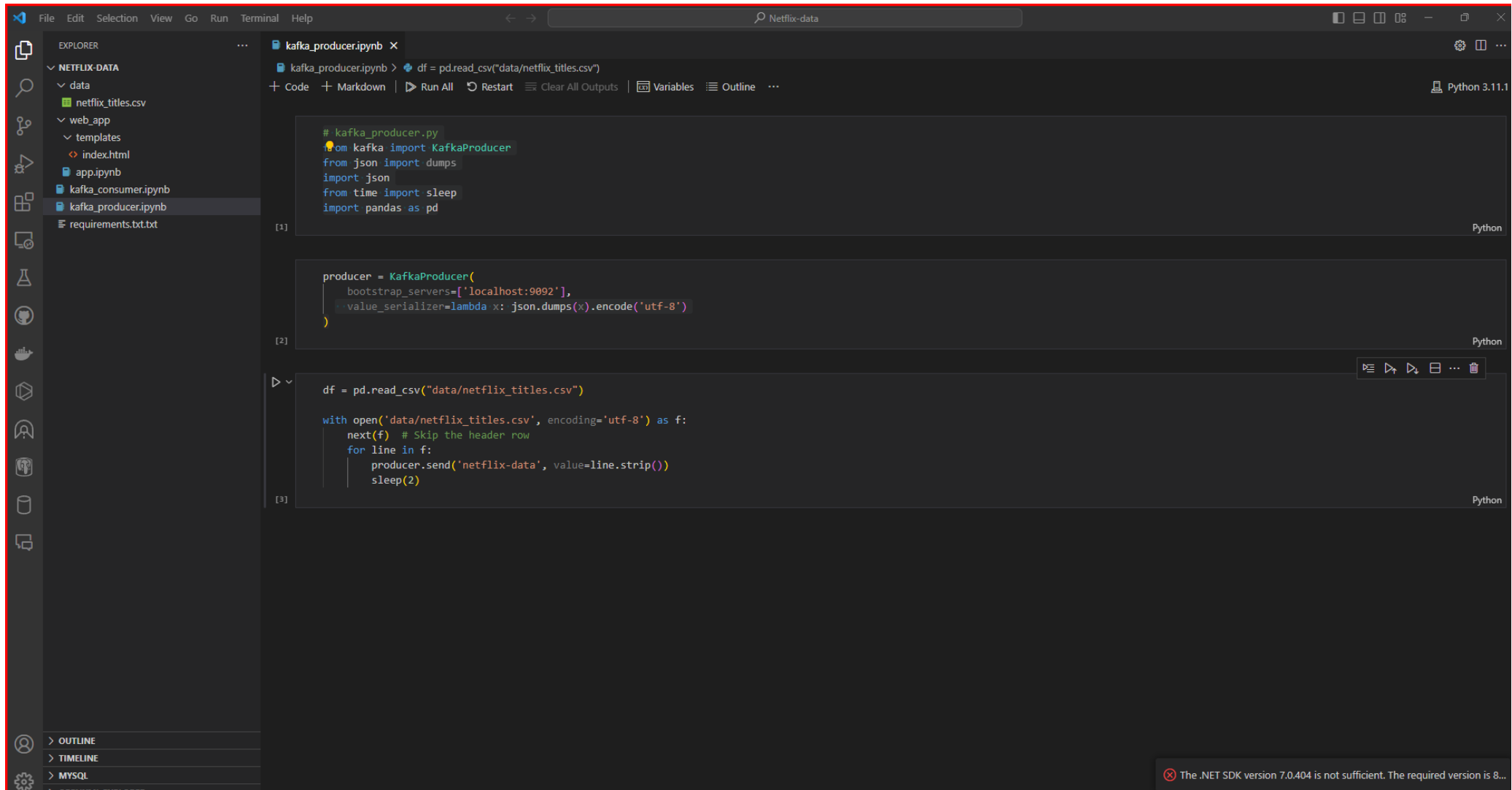
 Rules with source of 0.0.0.0/0 or ::/0 allow all IP addresses to access your instance. We recommend setting security group rules to allow access from known IP addresses only.

Cancel

Preview changes

Save rules

The screenshot displays the AWS Management Console interface. At the top, there's a navigation bar with the AWS logo, 'Services', a search bar, and a region dropdown set to 'N. Virginia'. Below this is a blue banner for 'Introducing myApplications on AWS' with a 'Create application' button. The main content area is titled 'Console Home' and features a 'Recently visited' list on the left including 'Billing and Cost Management', 'EC2', 'S3', 'IAM', 'IAM Identity Center', and 'AWS Supply Chain'. Below this are sections for 'Welcome to AWS' and 'AWS Health'. Two terminal windows are overlaid on the right side of the console. The top terminal window shows the command prompt 'C:\Windows\System32\cmd.e' and the execution of 'kafka-topics.bat --create --topic netflix-data --bootstrap-server localhost:9092 --replication-factor 1 --partitions 1', resulting in an error: 'Error while executing topic command : Topic 'netflix-data' already exists.' The bottom terminal window shows the execution of 'kafka-console-producer.bat --broker-list localhost:9092 --topic netflix-data'.



```
# kafka_producer.py
from kafka import KafkaProducer
from json import dumps
import json
from time import sleep
import pandas as pd

producer = KafkaProducer(
    bootstrap_servers=['localhost:9092'],
    value_serializer=lambda x: json.dumps(x).encode('utf-8')
)

df = pd.read_csv("data/netflix_titles.csv")

with open('data/netflix_titles.csv', encoding='utf-8') as f:
    next(f) # Skip the header row
    for line in f:
        producer.send('netflix-data', value=line.strip())
        sleep(2)
```

The screenshot displays a Jupyter Notebook environment with a file explorer on the left showing a project structure for 'NETFLIX-DATA'. The notebook contains three code cells:

```
[1] # kafka_producer.py
from kafka import KafkaProducer
from json import dumps
import json
from time import sleep
import pandas as pd
0.6s
```

```
[2] producer = KafkaProducer(
    bootstrap_servers=['localhost:9092'],
    value_serializer=lambda x: json.dumps(x).encode('utf-8')
)
0.1s
```

```
[3] df = pd.read_csv("data/netflix_titles.csv")

with open('data/netflix_titles.csv', encoding='utf-8') as f:
    next(f) # Skip the header row
    for line in f:
        producer.send('netflix-data', value=line.strip())
        sleep(2)
26.5s
```

Two Windows Command Prompt windows are open, showing the execution of Kafka commands:

```
C:\Windows\System32\cmd.e
Microsoft Windows [Version 10.0.22621.2715]
(c) Microsoft Corporation. All rights reserved.

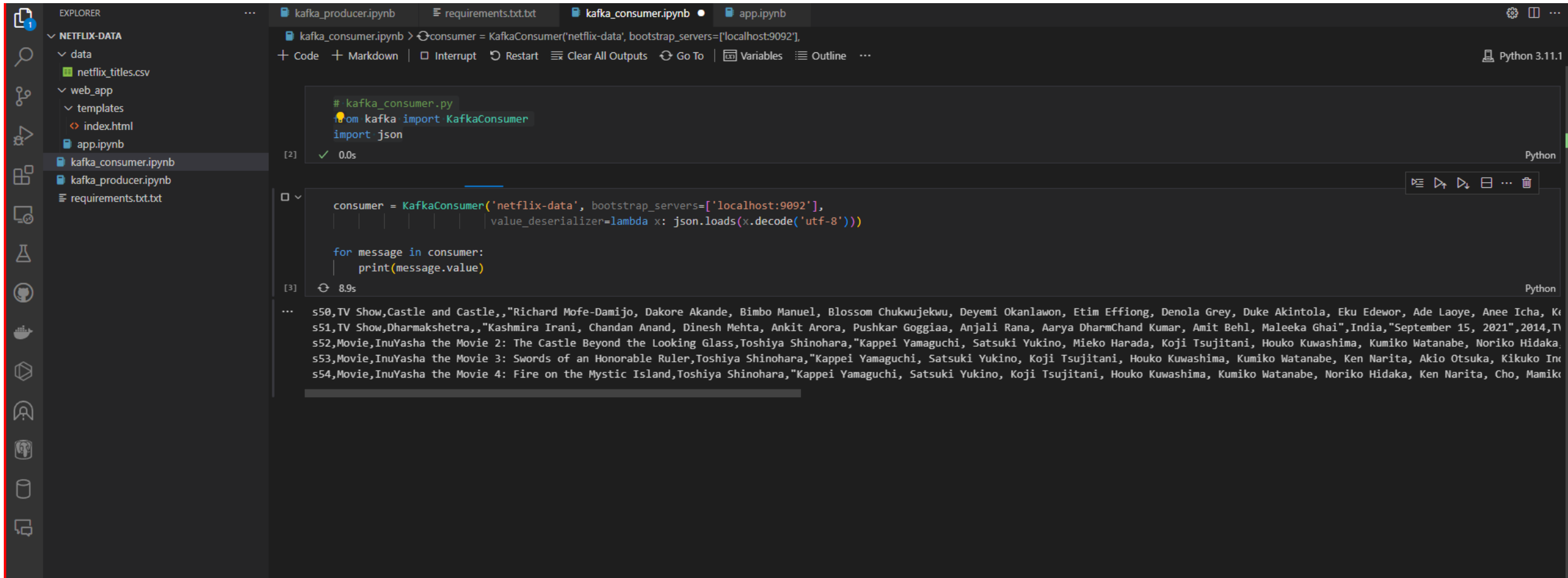
F:\kafka\bin\windows>kafka-topics.bat --create --topic netflix-data --bootstrap-server localhost:9092 --replication-factor 1 --partitions 1
Error while executing topic command : Topic 'netflix-data' already exists.
[2023-12-06 00:26:56,964] ERROR org.apache.kafka.common.errors.TopicExistsException: Topic 'netflix-data' already exists.
(kafka.admin.TopicCommand$)

F:\kafka\bin\windows>kafka-console-producer.bat --broker-list localhost:9092 --topic netflix-data

>jahir
>Emeka
>
```

```
C:\Windows\System32\cmd.e
h Republic\", \"September 23, 2021\", 2021, TV-MA, 127 min, \"Dramas, International Movies\", \"After most of her family is murdered in a terrorist bombing, a young woman is unknowingly lured into joining the very group that killed them.\"
\"s14, Movie, Confessions of an Invisible Girl, Bruno Garotti, \"Klara Castanho, Lucca Picon, J\u00falia Gomes, Marcus Bessa, Kiria Malheiros, Fernanda Concon, Gabriel Lima, Caio Cabral, Leonardo Cidade, Jade Cardozo\", \"September 22, 2021\", 2021, TV-PG, 91 min, \"Children & Family Movies, Comedies\", \"When the clever but socially-awkward Tet\u00e9 joins a new school, she'll do anything to fit in. But the queen bee among her classmates has other ideas.\"
```

The screenshot shows the AWS Console Home interface. The top navigation bar includes the AWS logo, 'Services' link, a search bar, and the region 'N. Virginia'. A blue banner at the top reads 'Introducing myApplications on AWS' with a 'Create application' button. The main content area is titled 'Console Home' and features a 'Recently visited' list with links to Billing and Cost Management, EC2, S3, IAM, IAM Identity Center, and AWS Supply Chain. Below this are sections for 'Welcome to AWS' (with a 'Getting started with AWS' link), 'AWS Health' (showing 0 open issues and 0 scheduled changes), and 'Cost Explorer' (showing current cost at \$0.00004 and forecasted month end costs at \$0.00003). Two terminal windows are overlaid on the right side of the console. The top terminal window shows the execution of 'kafka-topics.bat' to create a topic named 'netflix-data', which results in an error: 'Error while executing topic command : Topic 'netflix-data' already exists.' The bottom terminal window shows the execution of 'kafka-console-consumer.bat' to consume from the 'netflix-data' topic, with the output showing 'jahir' and 'Emeka'.



The screenshot displays a Jupyter Notebook environment with the following components:

- EXPLORER:** Shows a file tree with 'NETFLIX-DATA' containing 'data', 'web_app', and 'templates'. Other files include 'netflix_titles.csv', 'index.html', 'app.ipynb', 'kafka_consumer.ipynb', 'kafka_producer.ipynb', and 'requirements.txt.txt'.
- Code Editor:** Contains the following Python code in 'kafka_consumer.ipynb':

```
from kafka import KafkaConsumer
import json

consumer = KafkaConsumer('netflix-data', bootstrap_servers=['localhost:9092'],
                          value_deserializer=lambda x: json.loads(x.decode('utf-8')))

for message in consumer:
    print(message.value)
```
- Output:** Shows the execution of the code, resulting in a list of movie titles and details, such as 's50,TV Show,Castle and Castle,,Richard Mofe-Damijo, Dakore Akande, Bimbo Manuel, Blossom Chukwujekwu, Deyemi Okanlawon, Etim Effiong, Denola Grey, Duke Akintola, Eku Edewor, Ade Laoye, Anee Icha, K...'. The output is truncated with '...' at the end.

IMPLEMENT KAFKA CONSUMER

Consumer receiving data every 3 seconds from the producer

The screenshot displays a Jupyter Notebook environment with three files: `app.ipynb`, `kafka_consumer.ipynb`, and `kafka_producer.ipynb`. The `kafka_producer.ipynb` file contains the following Python code:

```
# kafka_producer.py
from kafka import KafkaProducer
import json
from time import sleep
import pandas as pd

producer = KafkaProducer(
    bootstrap_servers=['localhost:9092'],
    value_serializer=lambda x: json.dumps(x).encode('utf-8')
)
```

The `kafka_consumer.ipynb` file contains the following Python code:

```
df = pd.read_csv("data/netflix_titles.csv")

with open('data/netflix_titles.csv', encoding='utf-8') as f:
    next(f) # Skip the header row
    for line in f:
        # Convert each CSV line to a dictionary
        data = dict(zip(df.columns, line.strip().split(',')))
        producer.send('netflix-data', value=data)
        sleep(3)
```

The output of the first cell in `kafka_producer.ipynb` shows a successful execution with a time of 0.6s. The output of the second cell in `kafka_consumer.ipynb` shows a successful execution with a time of 13.7s.

Two Windows Command Prompt windows are also shown. The first window shows the execution of the following commands:

```
C:\Windows\System32\cmd.e
Microsoft Windows [Version 10.0.22621.2715]
(c) Microsoft Corporation. All rights reserved.

F:\kafka\bin\windows>kafka-topics.bat --create --topic netflix-data --bootstrap-server localhost:9092 --replication-factor 1 --partitions 1
Error while executing topic command : Topic 'netflix-data' already exists.
[2023-12-06 00:26:56,964] ERROR org.apache.kafka.common.errors.TopicExistsException: Topic 'netflix-data' already exists.
(kafka.admin.TopicCommand$)

F:\kafka\bin\windows>kafka-console-producer.bat --broker-list localhost:9092 --topic netflix-data

>jahir
>Emeka
>
```

The second window shows the output of the Kafka console producer, displaying a list of movie and TV show titles and descriptions, such as:

```
hiouche, "listed_in": " Noureddine Farihi", "description": " Geert Van Rampelberg"
{"show_id": "s4", "type": "TV Show", "title": "Jailbirds New Orleans", "director": "",
"cast": "", "country": "", "date_added": "\"September 24\"", "release_year": " 2021\"",
"rating": "2021", "duration": "TV-MA", "listed_in": "1 Season", "description": "\"Docu
ries"}
```

RETURN

EXPLORER

NETFLIX-DATA

data

netflix_titles.csv

web_app

templates

index.html

app.ipynb

kafka_consumer.ipynb

kafka_producer.ipynb

requirements.txt

OUTLINE

TIMELINE

MYSQL

ORIGINAL EXPLORER

kafka_producer.ipynb

kafka_consumer.ipynb

app.ipynb

web_app > app.ipynb > app = Flask(__name__)

+ Code

+ Markdown

Interrupt

Restart

Clear All Outputs

Go To

Variables

Outline

base (Python 3.11.4)

[1] ✓ 2.0s

Python

⌵

⏮

⏪

⏩

⏭

⌵

```

app = Flask(__name__)

# Load Netflix data
df = pd.read_csv("F:/BDAT SEM 1/Data Systems Architecture/Netflix-data/data/netflix_titles.csv")

# Extract descriptions for TF-IDF vectorization
df['description'] = df['description'].fillna('')
tfidf_vectorizer = TfidfVectorizer(stop_words='english')
tfidf_matrix = tfidf_vectorizer.fit_transform(df['description'])

# Global variable to store Kafka messages
kafka_messages = []

# Function to consume Kafka messages in a separate thread
def consume_kafka_messages():
    global kafka_messages
    consumer = KafkaConsumer('demo', bootstrap_servers=['localhost:9092'],
                             value_deserializer=lambda x: json.loads(x.decode('utf-8')))
    for message in consumer:
        kafka_messages.append(message.value)

# Start Kafka message consumption in a separate thread
kafka_thread = threading.Thread(target=consume_kafka_messages)
kafka_thread.start()

# Flask route to display recommendations
@app.route('/')
def index():
    global kafka_messages

    # Wait for 6 seconds to allow Kafka messages to be consumed
    time.sleep(6)

    # Extract descriptions from Kafka messages
    descriptions = [message['description'] for message in kafka_messages if 'description' in message]

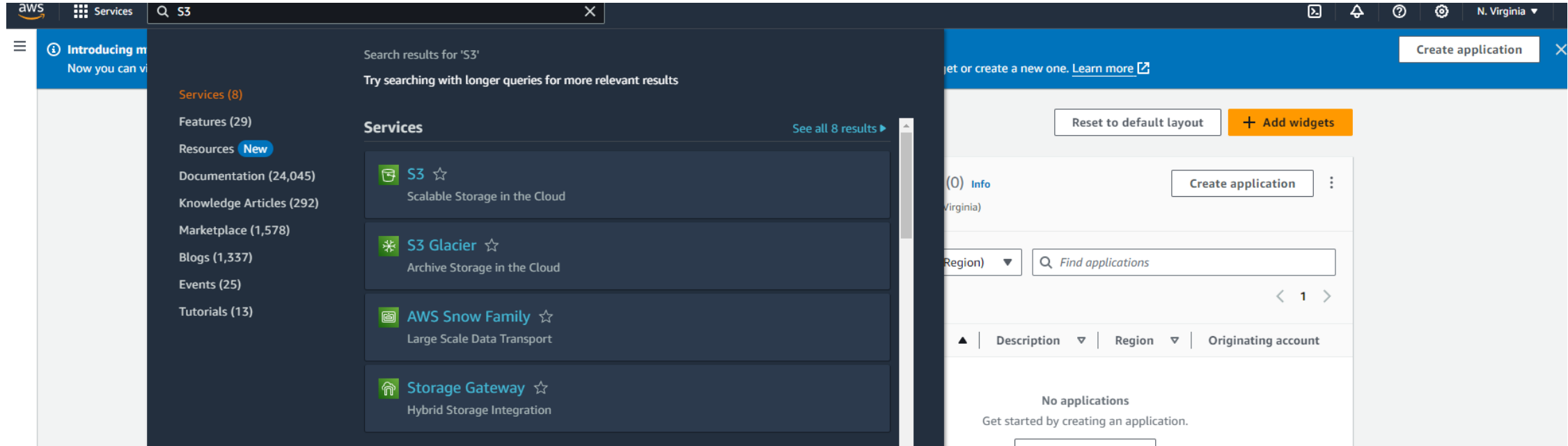
    # Transform descriptions using the TF-IDF vectorizer
    new_tfidf_matrix = tfidf_vectorizer.transform(descriptions)

    # Calculate cosine similarity between input and Netflix data
    cosine_sim = linear_kernel(new_tfidf_matrix, tfidf_matrix)

```

[RETURN](#)

S3 Bucket used to store Json data from Netflix





FLASK WEB INTERFACE



S3 Bucket Interface

aws

Services

Search

[Alt+S]

Global

Amazon S3

Buckets

Access Grants New

Access Points

Object Lambda Access Points

Multi-Region Access Points

Batch Operations

IAM Access Analyzer for S3

Block Public Access settings for this account

Storage Lens

Dashboards

Storage Lens groups New

AWS Organizations settings

Feature spotlight 7

► AWS Marketplace for S3

Amazon S3

► Account snapshot

Storage lens provides visibility into storage usage and activity trends. [Learn more](#)

View Storage Lens dashboard

General purpose buckets

Directory buckets

General purpose buckets (1) [Info](#)

Buckets are containers for data stored in S3. [Learn more](#)

Find buckets by name

< 1 > ⚙

	Name ▲	AWS Region ▼	Access ▼	Creation date ▼
○	netflix-data-jahir	Canada (Central) ca-central-1	<u>Bucket and objects not public</u>	December 4, 2023, 19:25:09 (UTC-05:00)

↻

Copy ARN

Empty

Delete

Create bucket

RETURN

FLASK WEB INTERFACE

Sending data to S3 Bucket in Json format

```
def upload_to_s3(data):
    # Specify your AWS S3 credentials and bucket name
    aws_access_key = 'AKIAI44QH8DHBVS7LH3MAOQ'
    aws_secret_key = 'C2X91234567890/dzqV1234567890XuvDyPHVNaa'
    bucket_name = 'my-bucket-hir'

    s3 = boto3.client('s3', aws_access_key_id=aws_access_key, aws_secret_access_key=aws_secret_key)

    try:
        # Generate a unique identifier for the filename
        unique_id = str(uuid.uuid4())

        # Convert data to JSON format
        json_data = json.dumps(data)

        # Upload data to S3
        s3.put_object(Bucket=bucket_name, Key=f'json_files/{unique_id}.json', Body=json_data)
        print(f"Data for ID {unique_id} uploaded to S3 successfully.")
    except NoCredentialsError:
        print("Credentials not available")

# Example consumer loop
for message in consumer:
    print(message.value)
    # Assuming message.value is a dictionary containing the data
    upload_to_s3(message.value)
```

```
{'show_id': 's107', 'type': 'TV Show', 'title': 'Bunk'd', 'director': '', 'cast': '"Peyton List", 'country': 'Karan Brar', 'date_added': 'Skai Jackson', 'release_year': 'Miranda May', 'rating': '4.5'}
Data for ID 26f3798e-f416-4794-83ae-cbc5e746439e uploaded to S3 successfully.
{'show_id': 's108', 'type': 'Movie', 'title': 'A Champion Heart', 'director': 'David de Vos', 'cast': '"Mandy Grace", 'country': 'David de Vos', 'date_added': 'Donna Rusch', 'release_year': 'Devar', 'rating': '4.5'}
Data for ID d5b46cf8-807c-4777-9779-42f65941b012 uploaded to S3 successfully.
{'show_id': 's109', 'type': 'TV Show', 'title': 'Dive Club', 'director': '', 'cast': '"Aubri Ibrag', 'country': 'Sana'a Shaik', 'date_added': 'Miah Madden', 'release_year': 'Mercy Cornwall', 'rating': '4.5'}
Data for ID 3bb9f01f-b450-417e-9854-b2a0970c3223 uploaded to S3 successfully.
{'show_id': 's110', 'type': 'TV Show', 'title': 'La casa de papel', 'director': '', 'cast': '"Úrsula Corberó', 'country': 'Itziar Ituño', 'date_added': 'Álvaro Morte', 'release_year': 'Paco Tous', 'rating': '4.5'}
Data for ID 04a90d20-e27b-4a2f-92c5-9d2730bb92de uploaded to S3 successfully.
{'show_id': 's111', 'type': 'TV Show', 'title': 'Money Heist: From Tokyo to Berlin', 'director': '"Luis Alfaro', 'cast': 'Javier Gómez Santander"', 'country': '', 'date_added': '', 'release_year': '2017', 'rating': '4.5'}
Data for ID ff01b3ee-0c31-4b4c-a63e-d806d0885576 uploaded to S3 successfully.
{'show_id': 's112', 'type': 'TV Show', 'title': 'Sharkdog', 'director': '', 'cast': '"Liam Mitchell', 'country': 'Dee Bradley Baker', 'date_added': 'Grey Griffin', 'release_year': 'Josh McDermitt', 'rating': '4.5'}
Data for ID 2d47ee85-cfad-416e-8992-17820126bc34 uploaded to S3 successfully.
{'show_id': 's113', 'type': 'Movie', 'title': 'Worth', 'director': 'Sara Colangelo', 'cast': '"Michael Keaton', 'country': 'Stanley Tucci', 'date_added': 'Amy Ryan', 'release_year': 'Shunori Ramar', 'rating': '4.5'}
Data for ID 695b5db9-c888-4440-b9c2-6d6ae3914b05 uploaded to S3 successfully.
```



```
if __name__ == '__main__':  
    try:  
        app.run(debug=False)  
    except Exception as e:  
        print("An error occurred:", str(e))
```

[2]

...

```
* Serving Flask app '__main__'  
* Debug mode: off  
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.  
* Running on http://127.0.0.1:5000  
Press CTRL+C to quit  
127.0.0.1 - - [06/Dec/2023 00:46:26] "GET / HTTP/1.1" 200 -  
127.0.0.1 - - [06/Dec/2023 00:46:26] "GET /static/styles/style.css HTTP/1.1" 200 -  
127.0.0.1 - - [06/Dec/2023 00:46:26] "GET /favicon.ico HTTP/1.1" 404 -  
127.0.0.1 - - [06/Dec/2023 00:47:10] "GET / HTTP/1.1" 200 -  
127.0.0.1 - - [06/Dec/2023 00:47:10] "GET /static/styles/style.css HTTP/1.1" 304 -  
127.0.0.1 - - [06/Dec/2023 00:52:06] "GET / HTTP/1.1" 200 -  
127.0.0.1 - - [06/Dec/2023 00:52:06] "GET /static/styles/style.css HTTP/1.1" 304 -  
127.0.0.1 - - [06/Dec/2023 00:52:28] "GET / HTTP/1.1" 200 -  
127.0.0.1 - - [06/Dec/2023 00:52:28] "GET /static/styles/style.css HTTP/1.1" 304 -
```



FLASK WEB INTERFACE



Console Home | Console Home xNetflix Data Dashboard x+

127.0.0.1:5000

Apps Gmail Maps Python for Every... General (AIDI Stud... React - The Compl... iCloud Mail fullstack-course4/L... YouTube notes - Google Drive Dive into Python [...] Theory of Computa... Dashboard - Chhot... All Bookmarks

Netflix Data Dashboard

Latest Shows and Movies

Jaws 3 Movie

Director: Joe Alves

Country: Bess Armstrong

Release Year: Louis Gossett Jr.

Duration: Lea Thompson

Description: Dan Blasko

Cast: "Dennis Quaid"

Date Added: Simon MacCorkindale

Rating: John Putch

Genres: P.H. Moriarty

Jaws: The Revenge Movie

Director: Joseph Sargent

Country: Lance Guest

Release Year: Karen Young

Duration: Judith Barsi

Description: Lynn Whitfield"

Cast: "Lorraine Gary"

Date Added: Mario Van Peebles

Rating: Michael Caine

Genres: Mitchell Anderson

My Heroes Were Cowboys Movie

Director: Tyler Greco

Country:

Release Year: 2021"

Cast:

Date Added: "September 16

Rating: 2021

★ Top 5 Recommendations

Check out more from director Joe Alves: Jaws 3

Featuring "Dennis Quaid, also in: Jaws 3

Check out more from director Joseph Sargent: Jaws: The Revenge

Featuring "Lorraine Gary, also in: Jaws: The Revenge

Check out more from director Tyler Greco: My Heroes Were Cowboys

RETURN

THANK YOU

RETURN