**KAFKA AND ITS USAGE AND INTEGRATION IN PEGA**  
  
  
Basically when we have a target and source system Data access between both of the systems would be easy…   
But when it becomes 20-30 systems, then sourcing data between target and source systems would become very difficult…  
  
So To avoid that KAFKA service has been introduced…  
What it will do is all the data from the source system would be stored in the Kafka cluster and target systems will source data directly from the Kafka cluster…

It is created by LinkedIn and is now an Open source project mainly maintained by Confluent and actively maintained by Apache.

**Examples---**

**Netflix** uses Kafka to apply recommendations in real time while you’re watching TV shows

**Uber** uses Kafka to gather user, taxi, and trip data in real-time to compute and forecast demand and surge pricing in real-time.

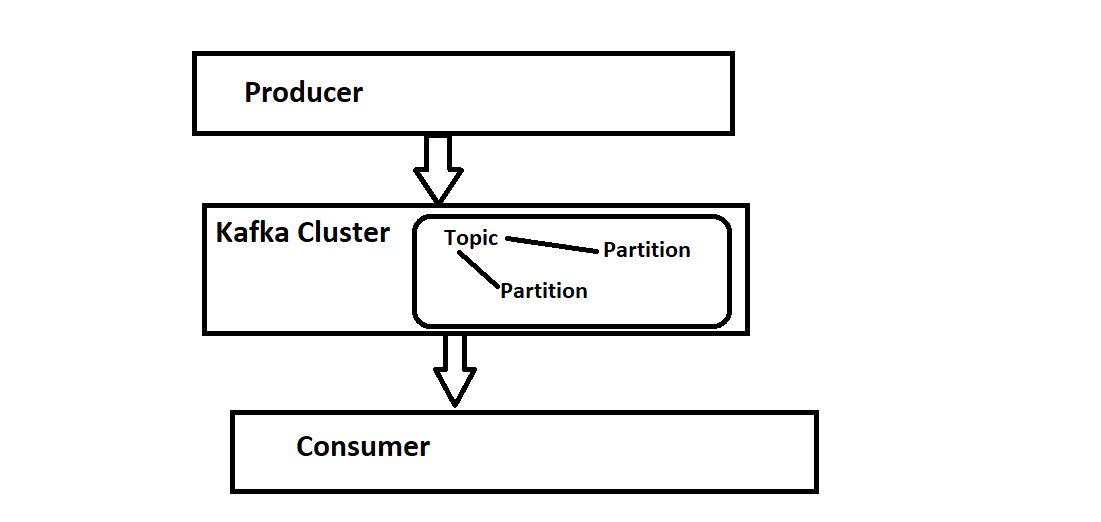
**LinkedIn** uses Kafka to prevent spam, and collect user interactions to make connection better recommendations in real time…

In Kafka, we have Topics that are nothing but (a particular stream of data) similar to a database…

And topics are split into partitions.

Each partition is ordered.. and each message within a partition gets an incremental id, called offset.

Once data is written to a partition, it can’t be changed.. and a key must be provided.

Data is assigned randomly to a partition unless a key is provided.  
  


Let us assume each producer is an application and its consumer is one application

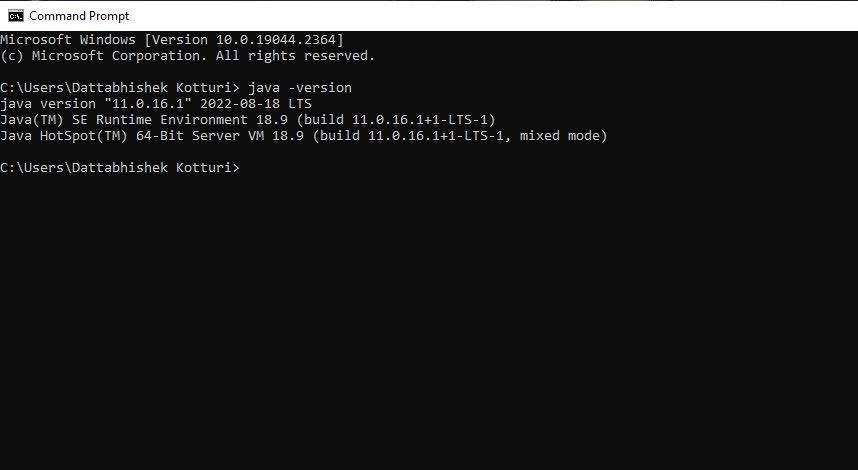
And the Kafka cluster will act as a broker between all the applications.

Let us assume I am the consumer and I want to get data from the application. So I would connect to all the producers individually.

So the Kafka cluster will act as a broker and the producer will push information to the Kafka cluster.. and consumer will subscribe and we will use the data

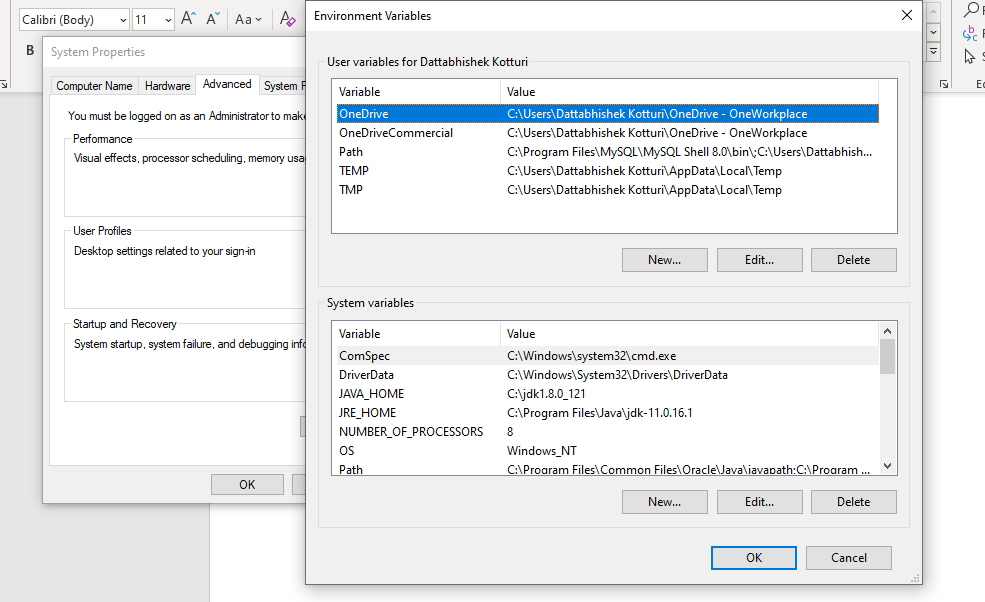
**EXAMPLE**- Let us say an end user is a consumer and producer as the udemy which has a playlist of Pega and java and python. So here the playlists would be pushed to the Kafka cluster which is a broker and then we would as end users subscribe and watch those videos.   
  
Now, this Kafka is maintained and supplied by Apache.  
  
So before going to Pega and writing and showing data from Kafka let us see how we run Kafka on the local system…

First, we need to have JAVA on our system. I am not going into detail about java installation. Just you need to have version 8 or higher… So I have java 11 installed on my system.

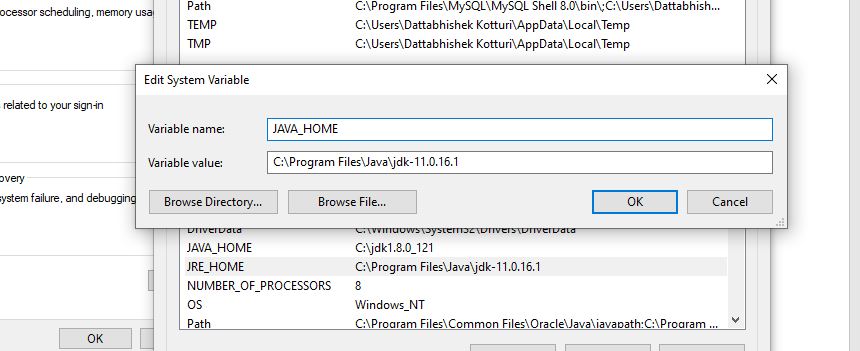


And then you need to set up environmental variables.

Control Panel > System Properties > Environment Variables

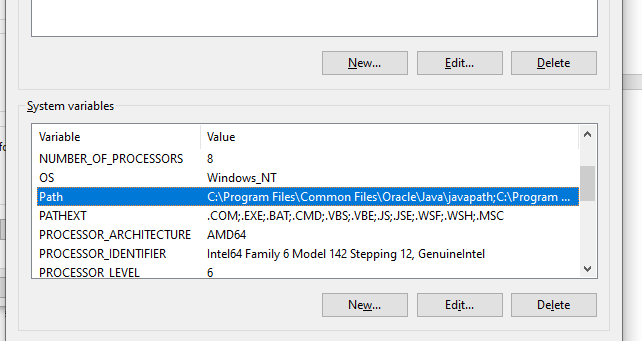


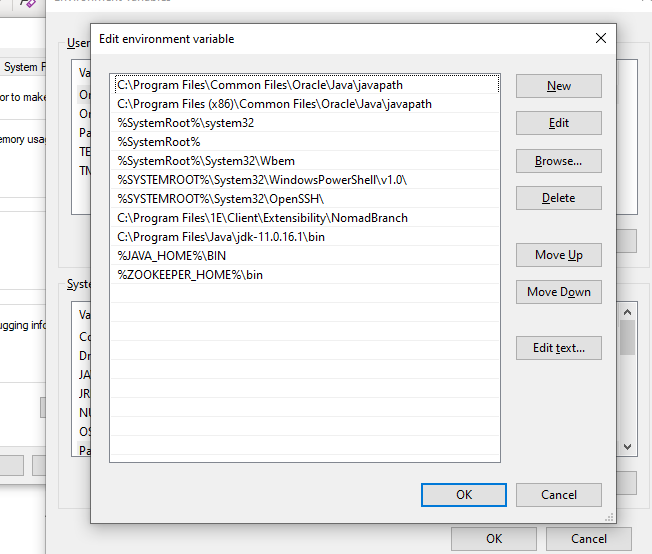
And inside it create as JAVA\_HOME and give the java directory where you have your JDK version



And save it and after that again check for the path and inside it edit and add

“%JAVA\_HOME%BIN”





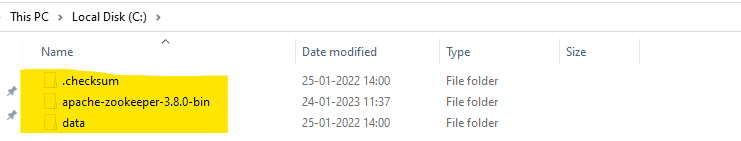
And after that, we have to install zookeeper and Kafka

So let’s install a zookeeper

Navigate to the <https://zookeeper.apache.org/releases.html> website and install zookeeper

The current version is Zookeeper 3.8.0

And after installation extract it you would be finding a folder like this

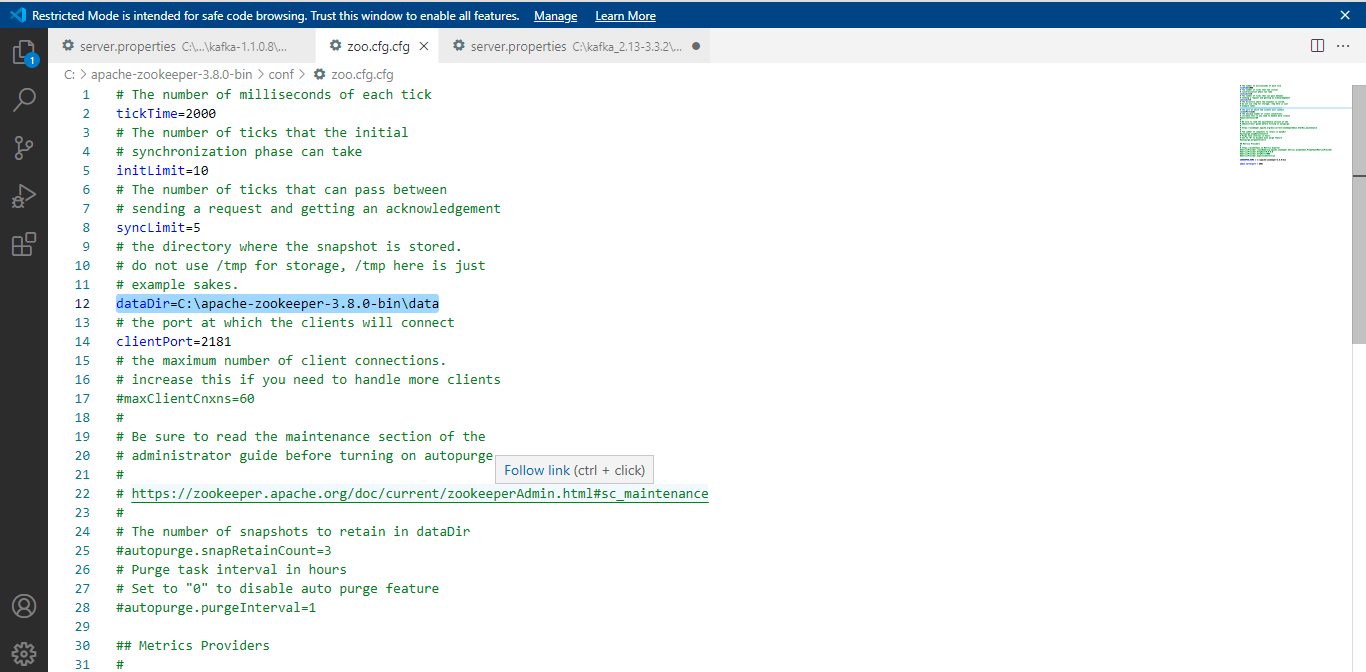


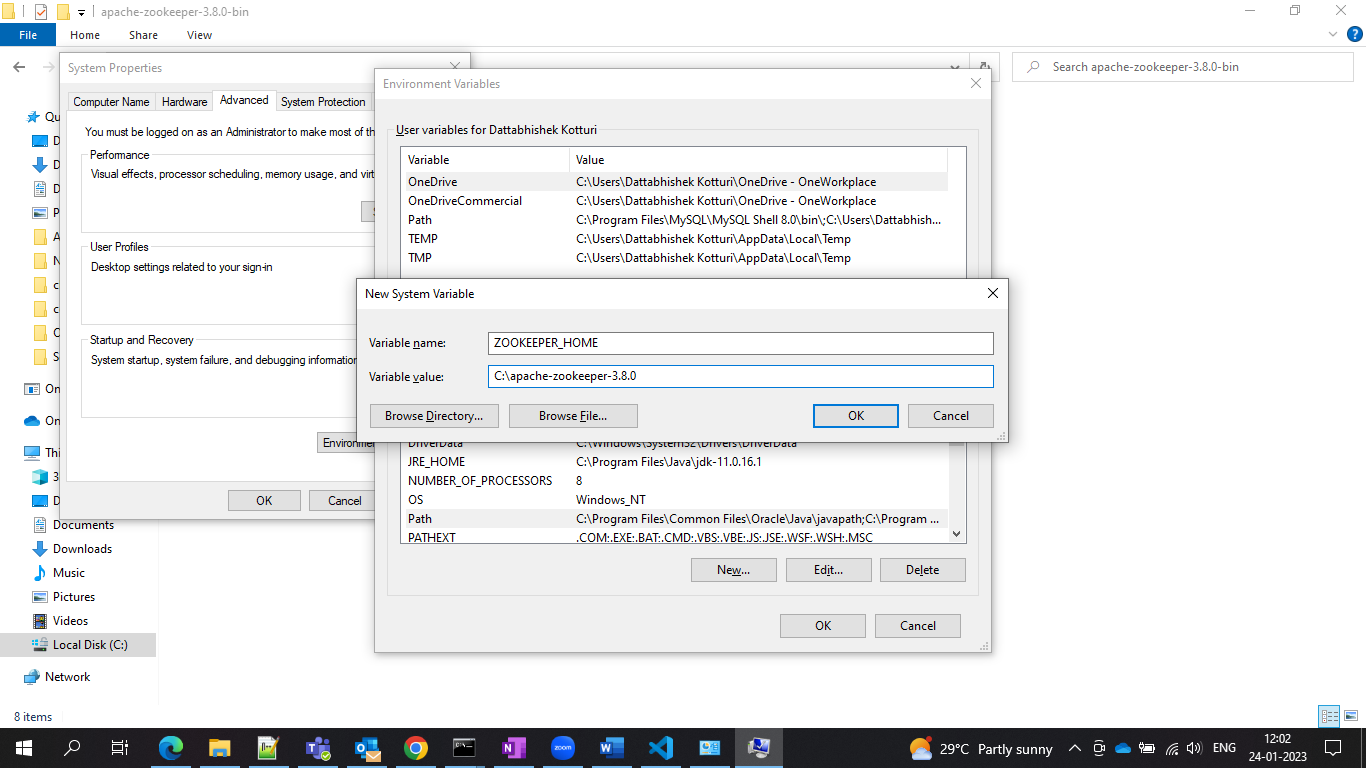
And inside it navigate to the config directory.. and inside that, you would find the file “zoo\_sample.cfg” and you can edit it to “zoo. cfg”

And now you can please open that file I am using VS code to open it

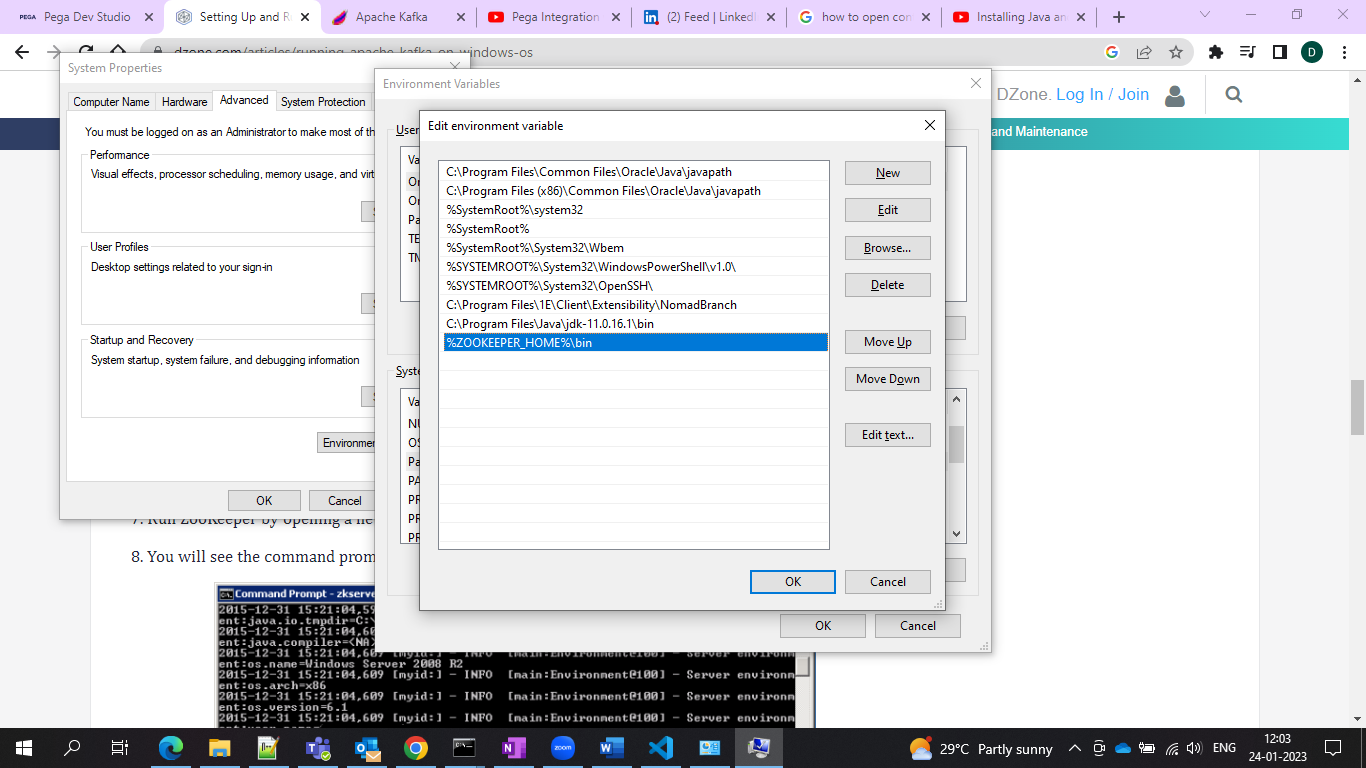
In that, you can find and edit

dataDir=/tmp/zookeeper to “dataDir=C:\apache-zookeeper-3.8.0-bin\data”

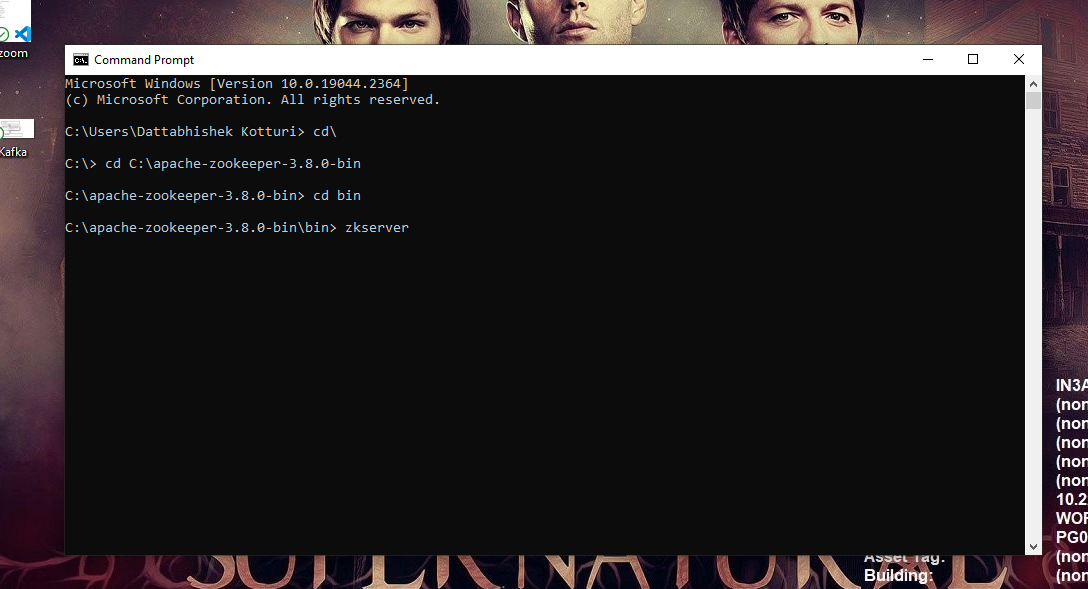


and after that again we need to go to the system environment variables and we need to add again the variables and we can 

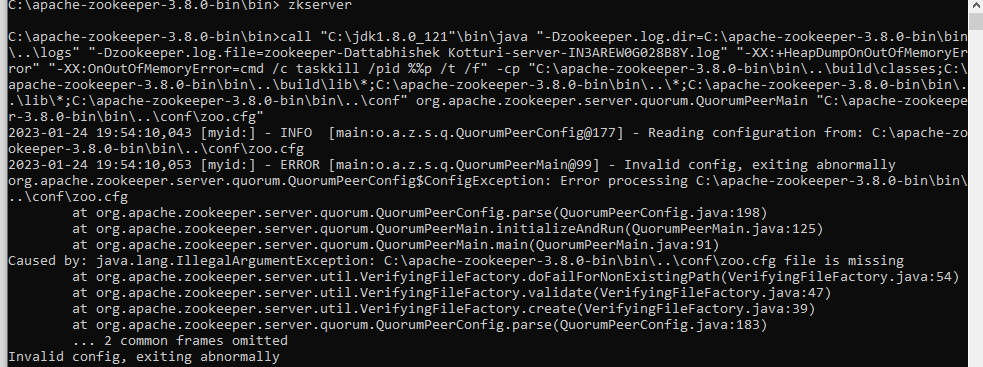
and after this, you can please find “path” and add the environment variable as shown in the screenshot.



And then you can check for the zookeeper default port as 2181 and you can run the command prompt and run the zookeeper server



Navigate to the apache directory and type server to start the zookeeper server



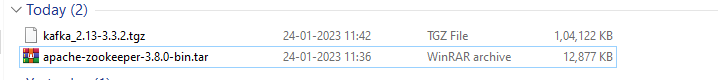
So the server will run and the server is started.. and the binding port is 2181

And now let us jump to Kafka and install it and configure it and make it running

You can install Kafka.

You can install it from <https://kafka.apache.org/downloads.html>

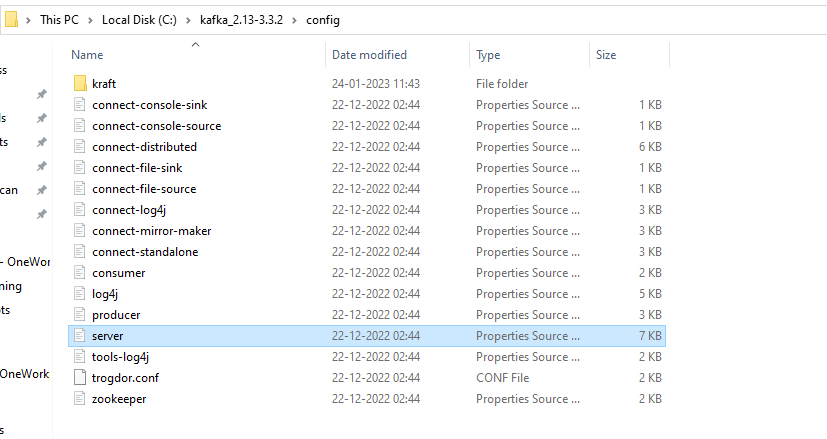
And now after installing it you would find a file like this…



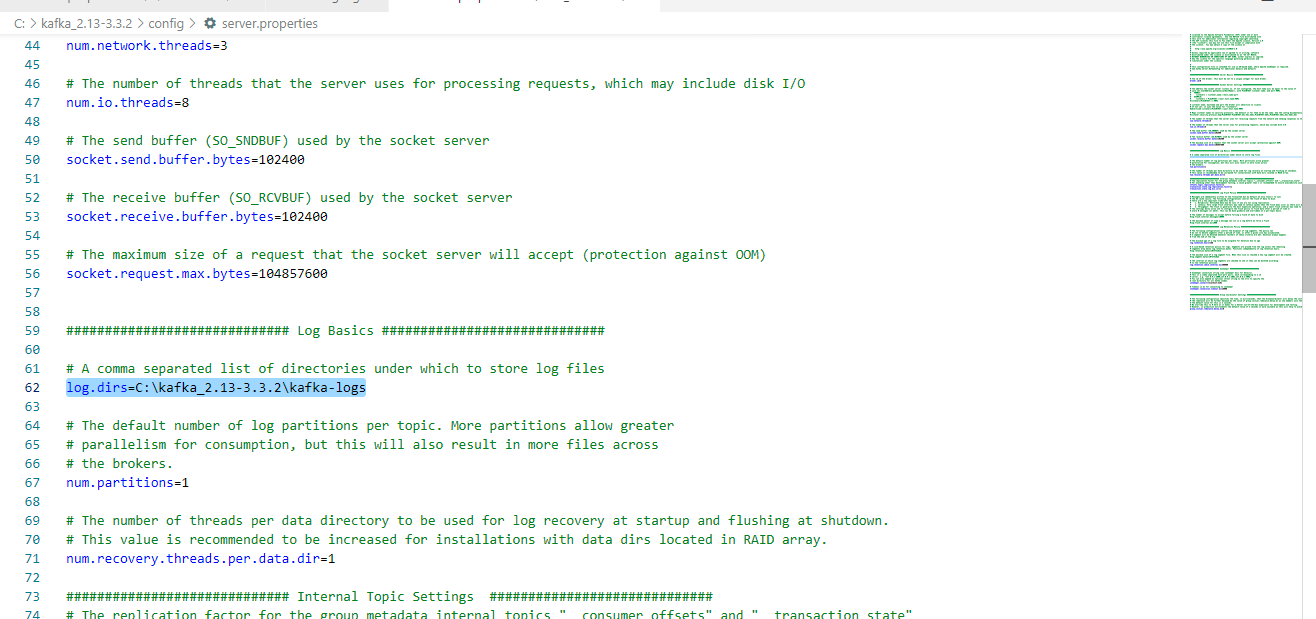
And now extract the Kafka----.tgz file (tgz means TAR Archive file that has been compressed using Gnu Zip (gzip) software…

And after the extraction, you would find a folder like this

And now please navigate to the config folder and open the server properties source file.

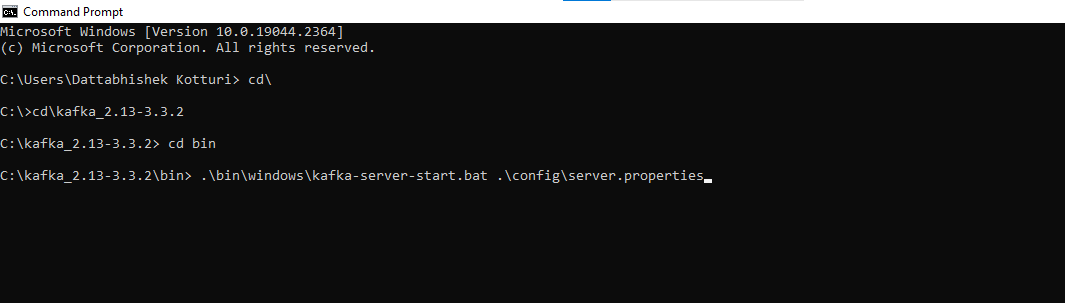


And then open it and find “log.dirs=/tmp/kafka-logs” and edit to “log.dir= C:\kafka\_2.11-0.9.0.0\kafka-logs”

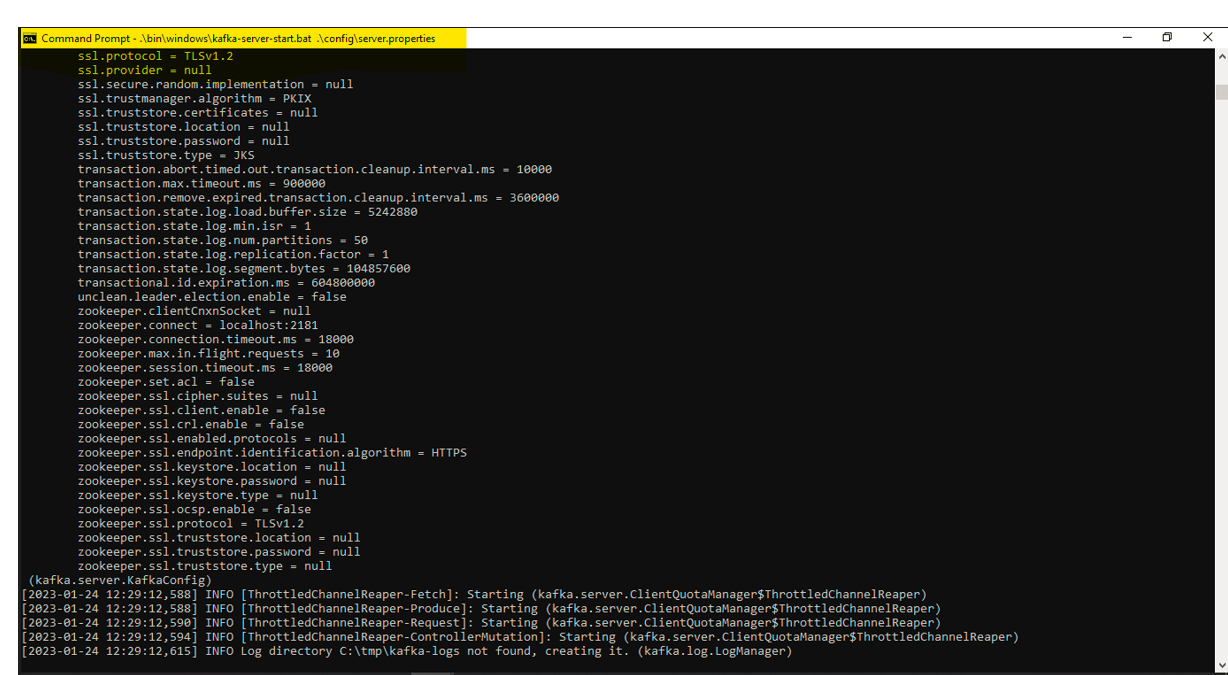


And now we should run the Kafka server and make sure that your zookeeper server is up and running

And now open the command prompt and navigate to the Kafka directory and type in the following

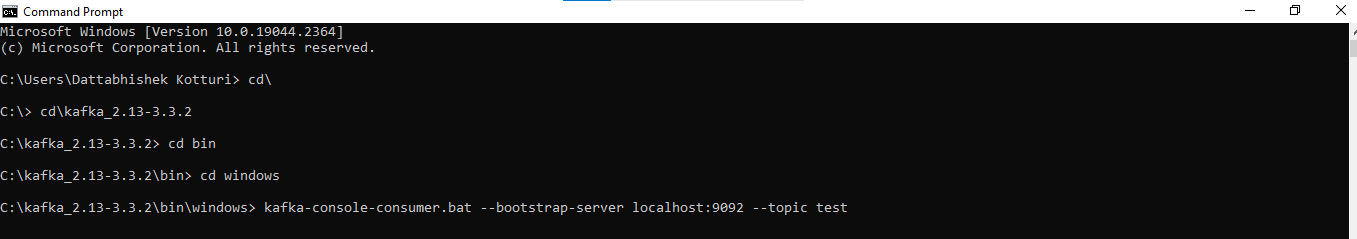


And just press enter and your Kafka server will start running.



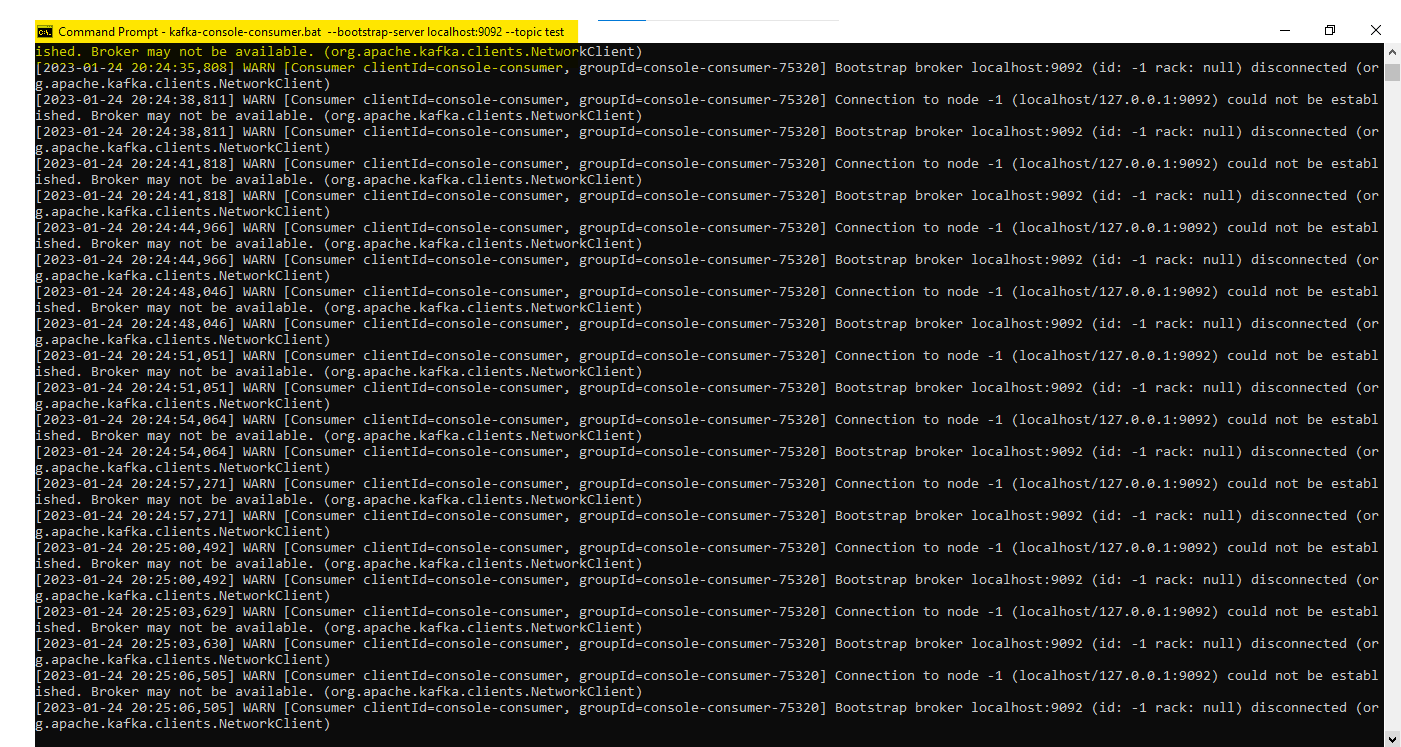
And now the Kafka server is up and running…

Now we have to create one Topic we need to again navigate to the command prompt and go to the Kafka directory\ bin\ windows and type the following

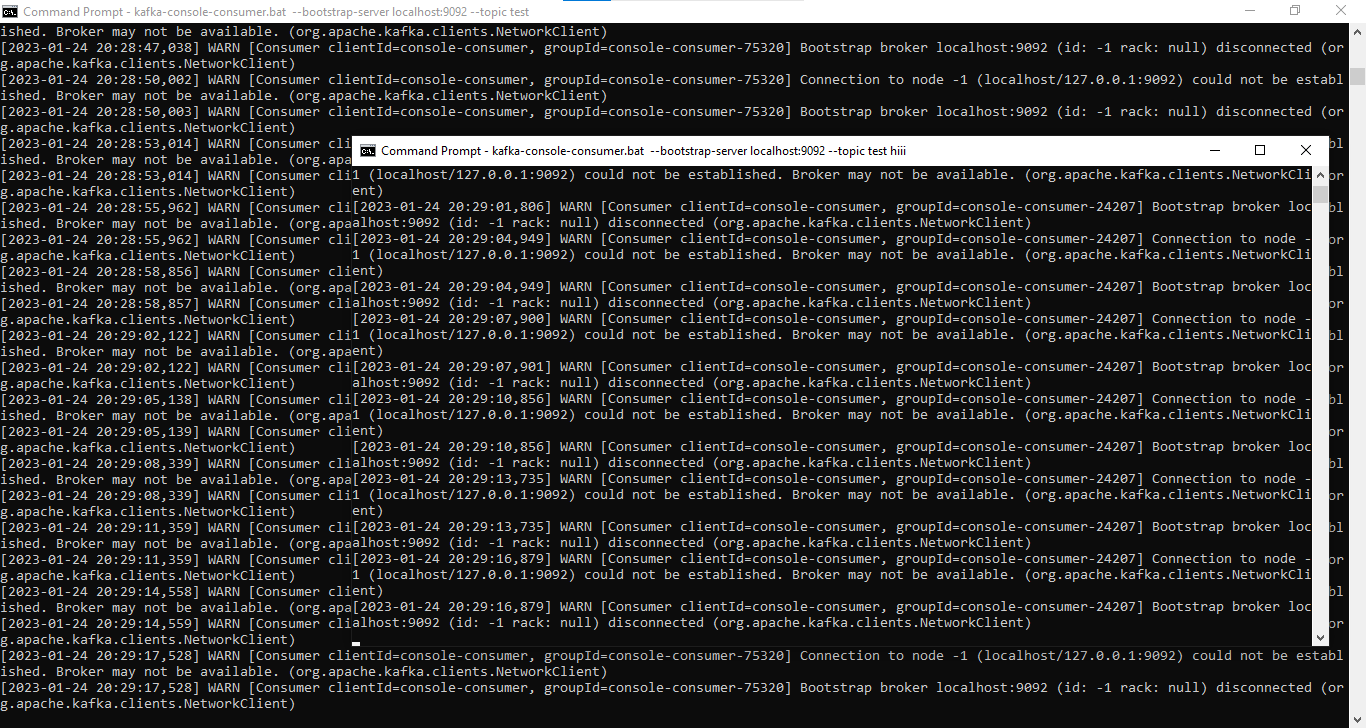


And start executing it

And it will create a topic

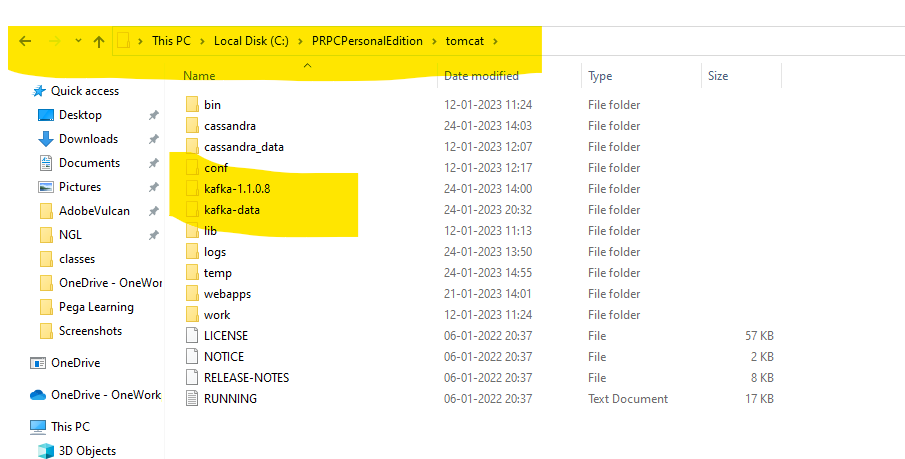


And then we would see two command prompts one with the consumer and one with the producer now if you can send and receive messages your Kafka configuration is done.

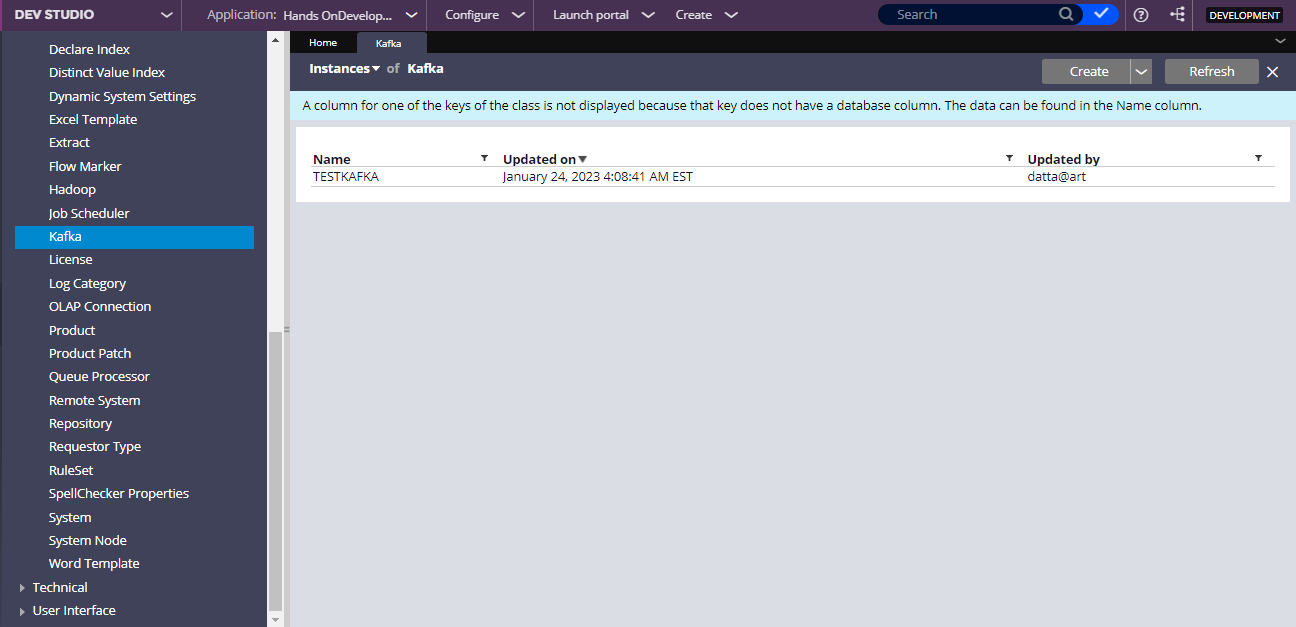


So now let’s get into Pega and test our Kafka through the data set…

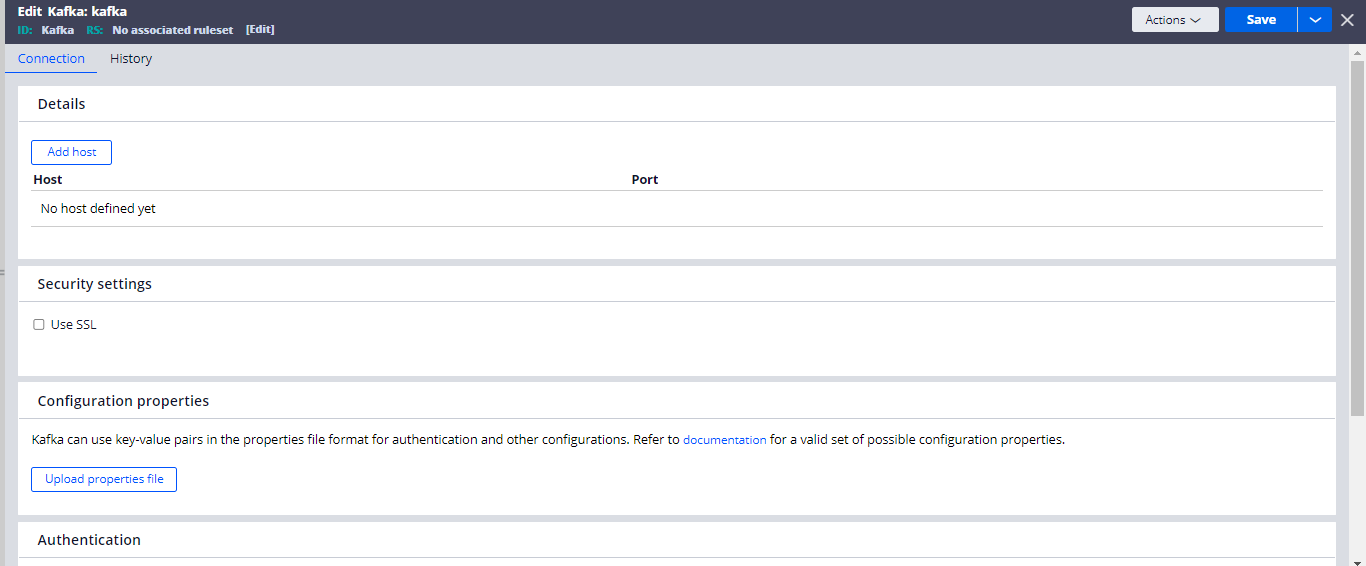
So for your information, installed install our Pega Personal Edition, we already have a Kafka database…



And in our developer studio, we would be finding the Kafka under SysAdmin



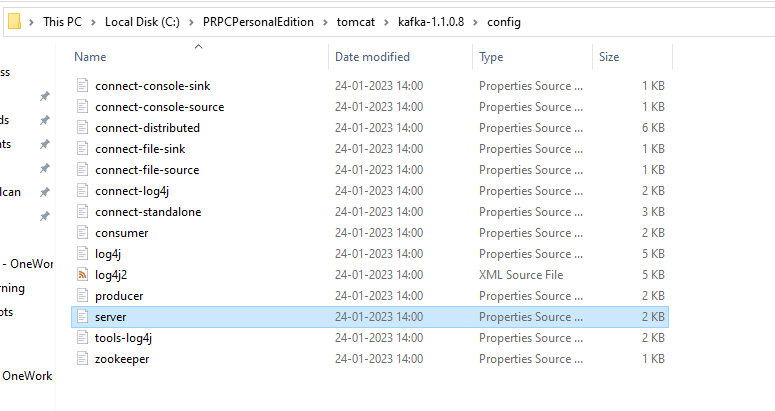
And now let us create a Kafka instance now



And for obtaining the host and port and where the Kafka broker is running we should navigate to the

C:\PRPCPersonalEdition\tomcat\kafka-1.1.0.8\config and under that open the server

Properties source file



And after opening you would find listeners=PLAINTEXT\://10.227.33.90\:9092

So here the host is 10.227.33.90

And the port is 9092

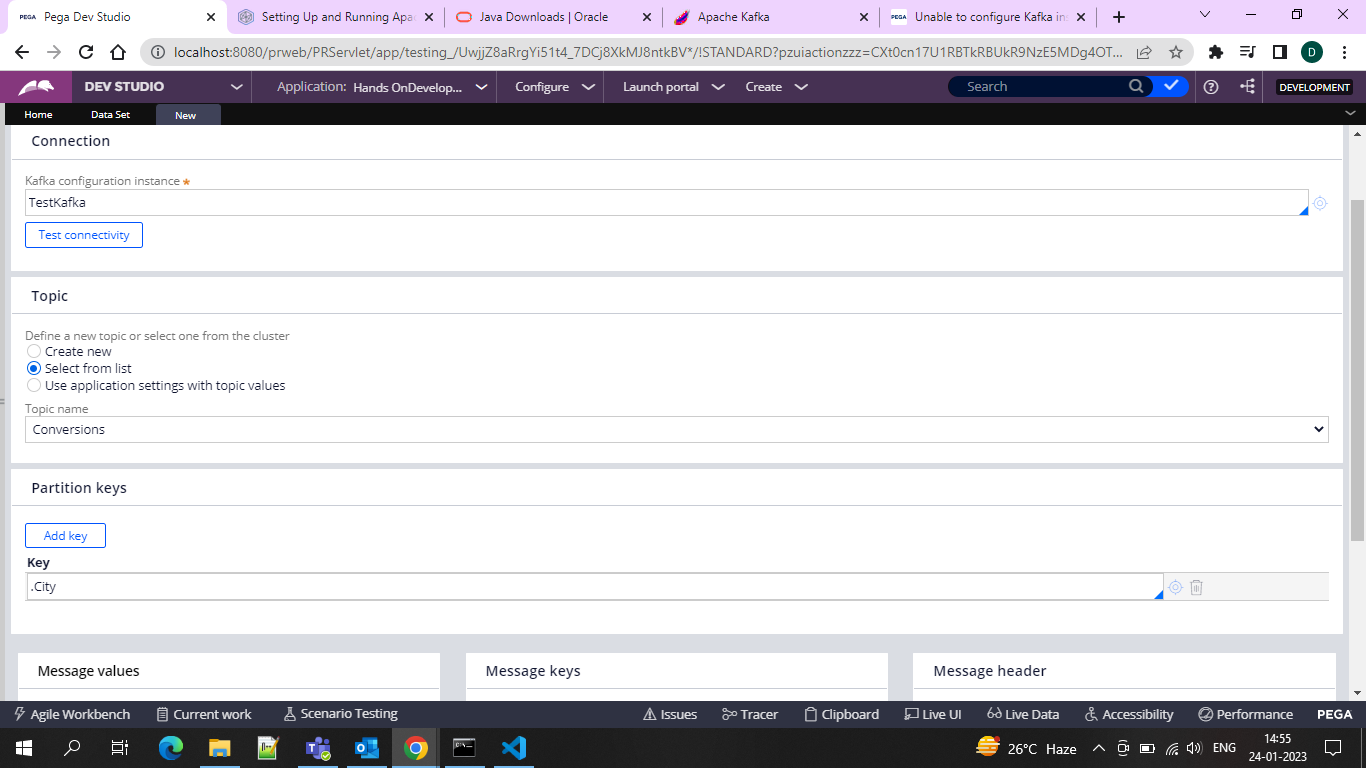
and after that save the rule and test the connectivity.

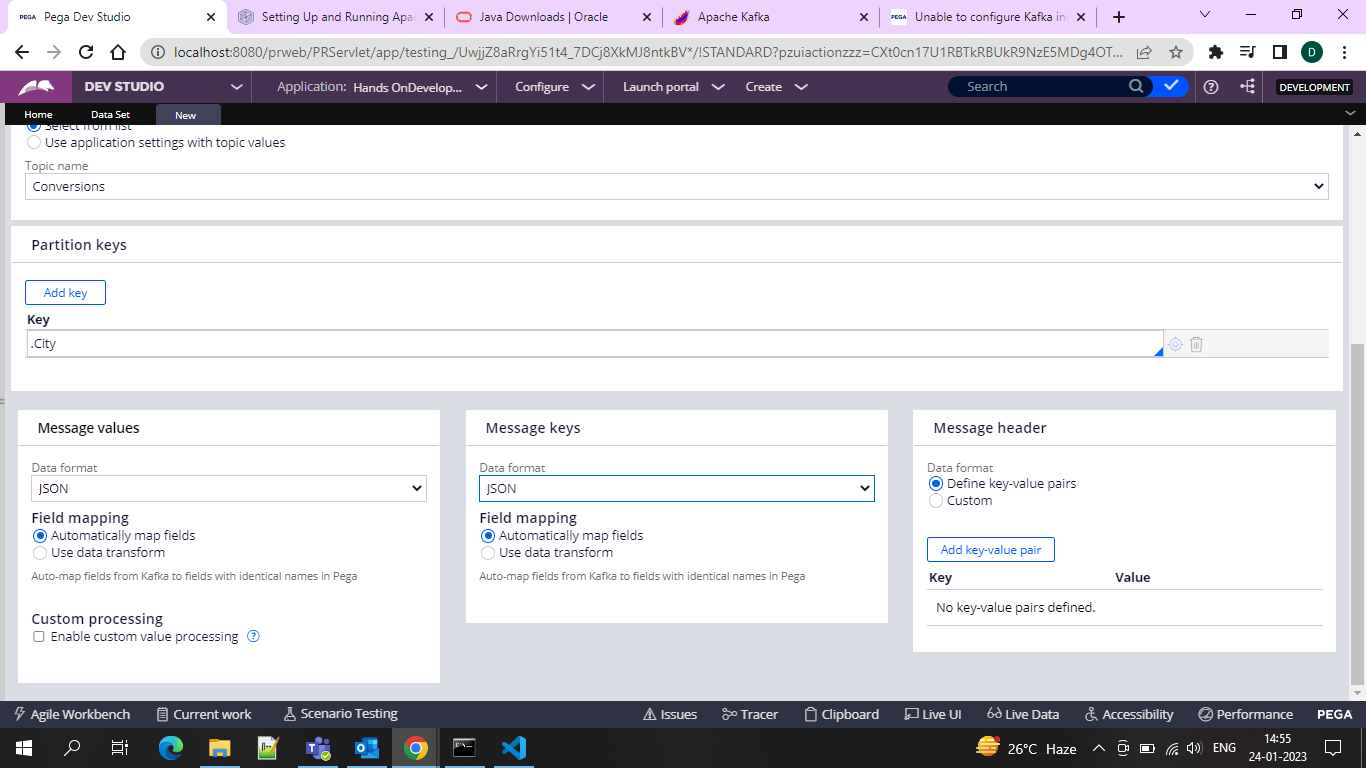
Now I want to publish the messages through the dataset.

So let’s create one data set of type ‘Kafka”

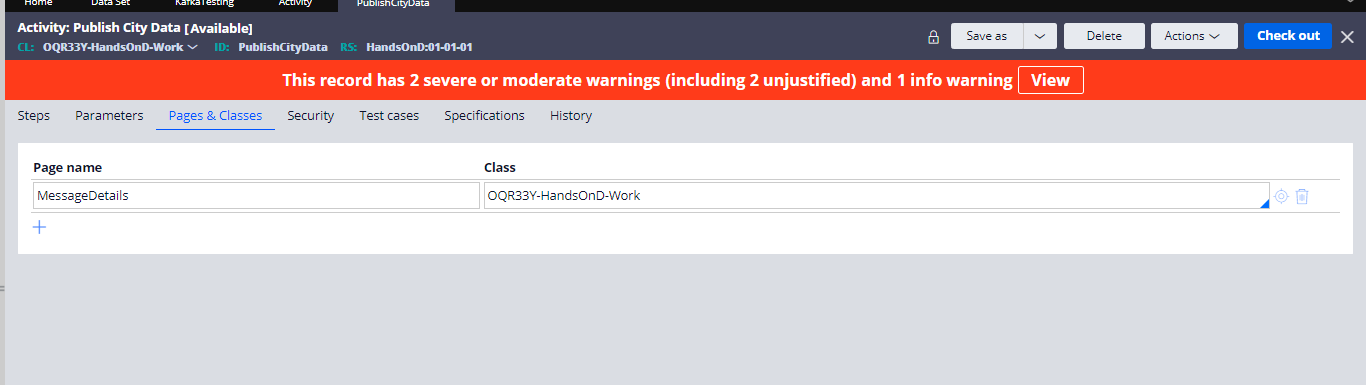
Make sure that the Kafka instance name and dataset name are the same

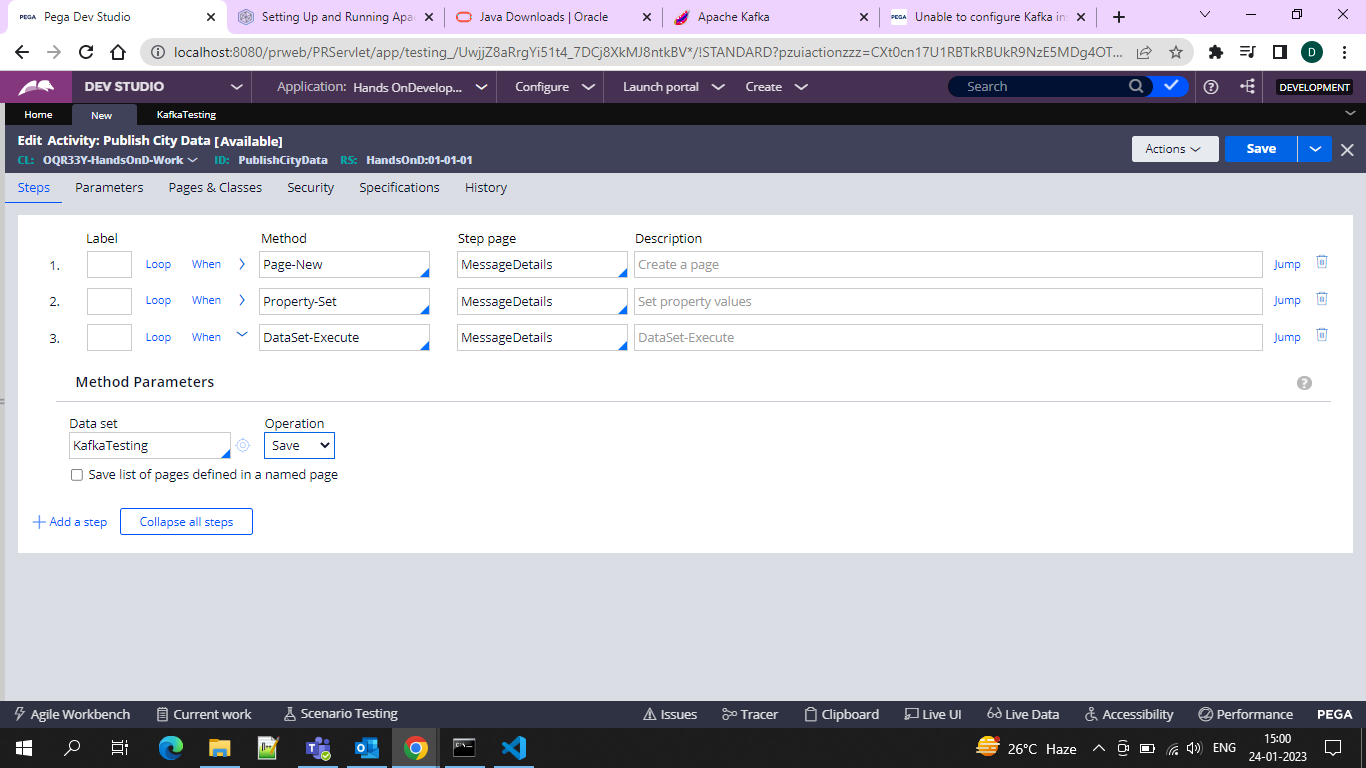
Now give your Kafka instance which you created and under the topic, I am choosing the existing one if wanted can create a new one.



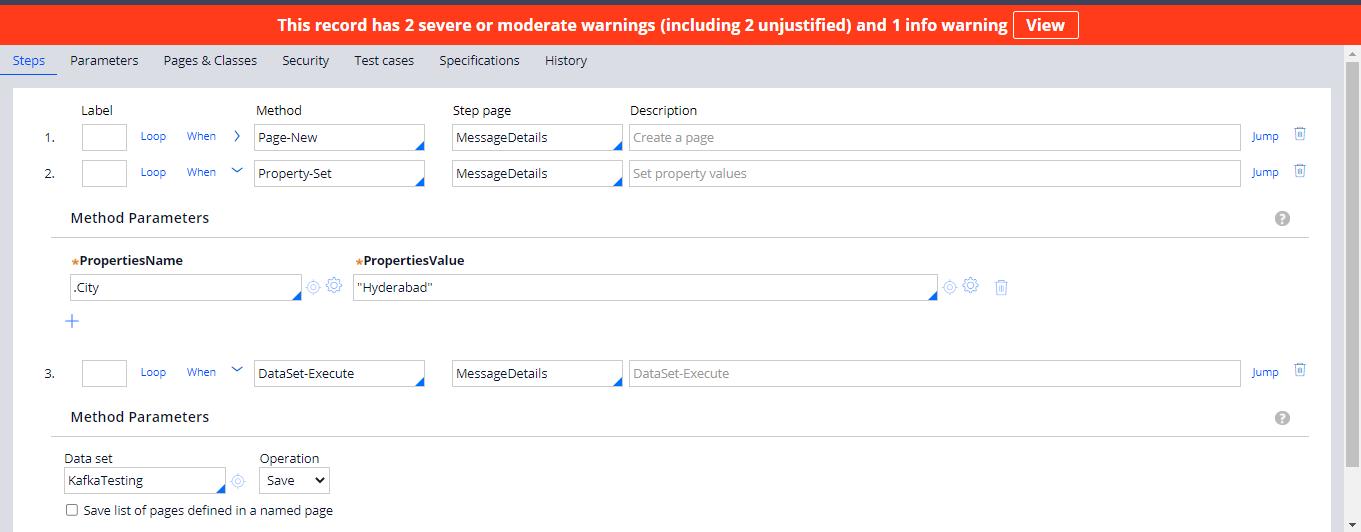
And in the message values you can select JSON and Message Keys also use JSON and message header select Define Key-Value Pairs 

And then use activity to run the dataset.  
  
So configure the activity as below



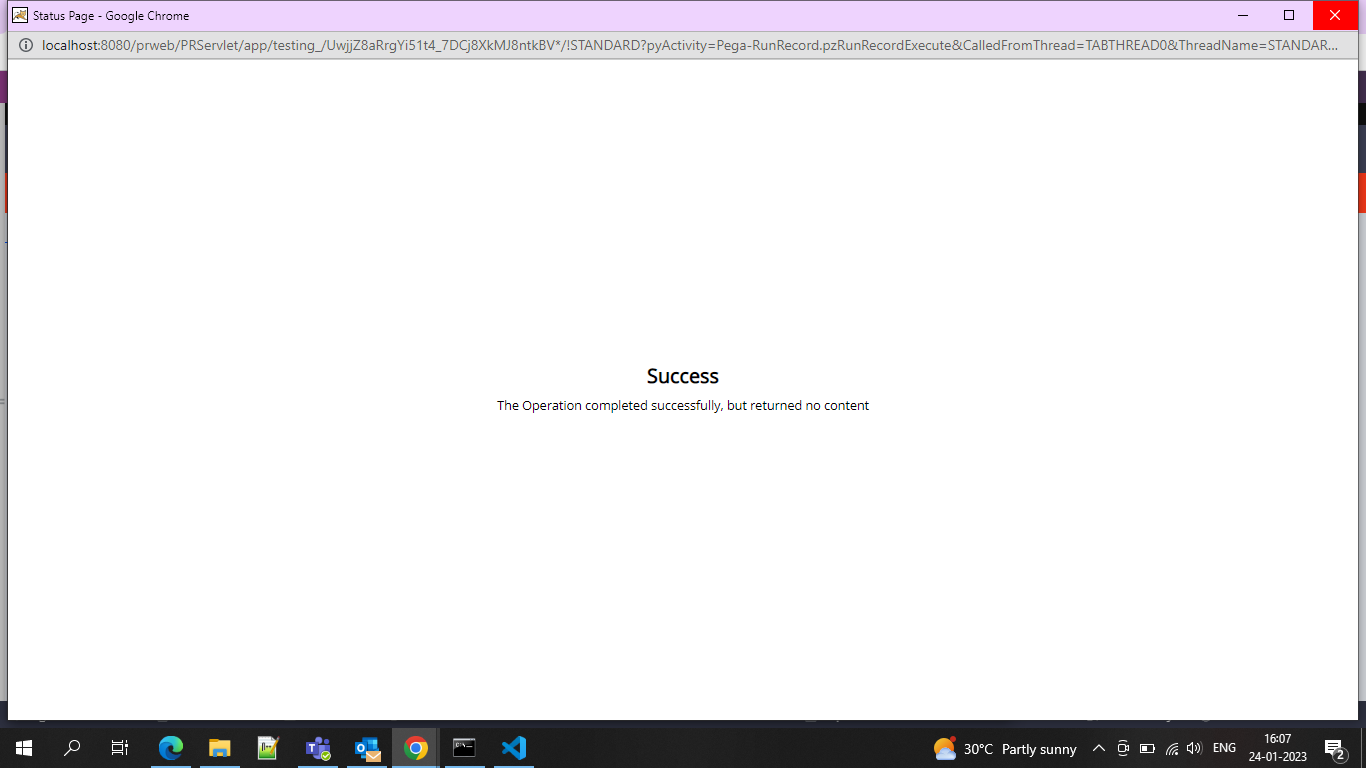


And then in the property set, you can define what you want to publish



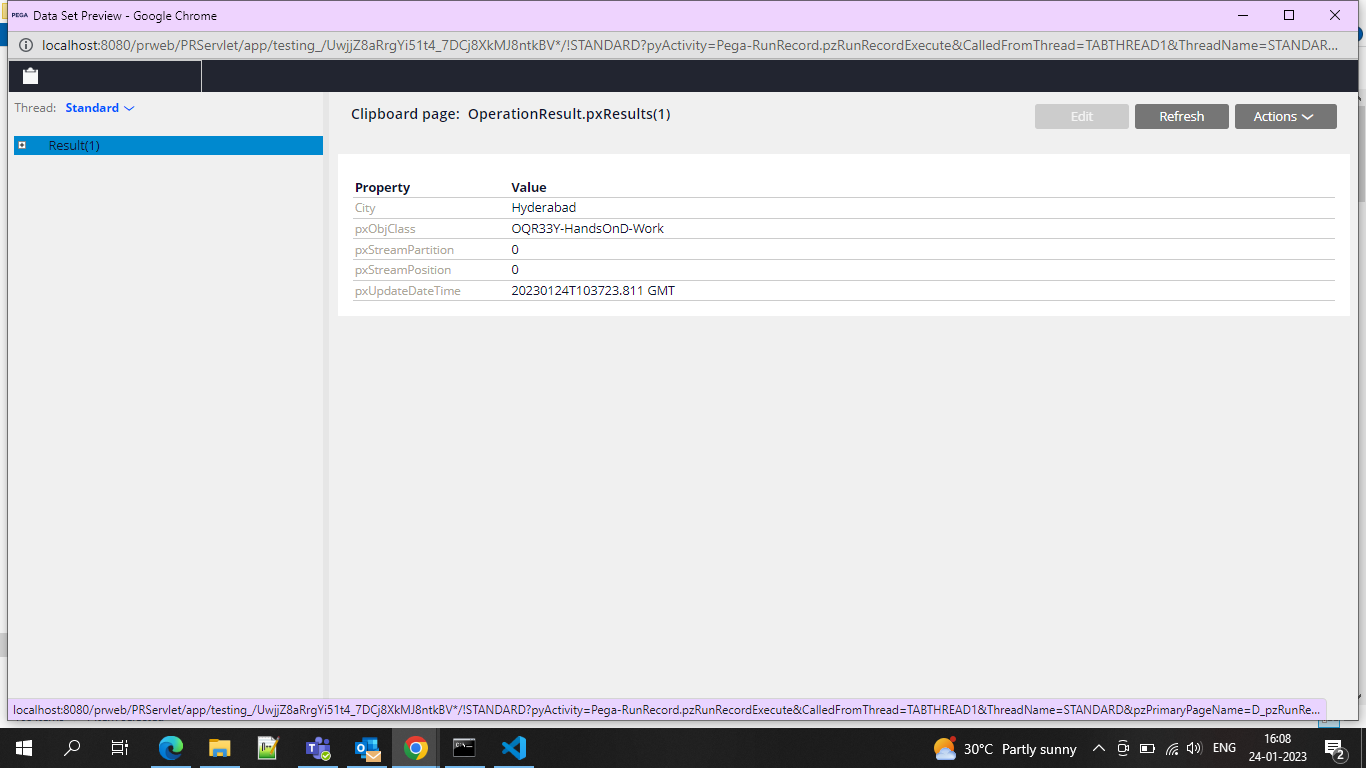
And in the dataset-execute method, you can use the start operation….

And now save and run the activity



And now run the data set and run and the record is published….

BOOM



So hence we can publish messages…