

# Kafka Installation

Prepared by: Dev Jethva (23MDS003)

Branch: MTech. (Data Science)

## ❖ Prerequisites

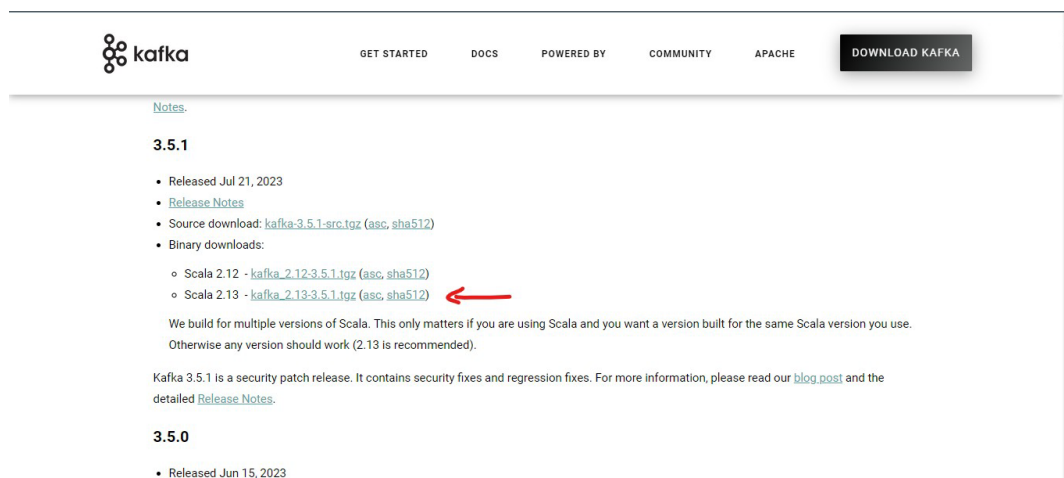
1. Java 8 development Kit (JDK)
2. 7zip

## ❖ Download Kafka file

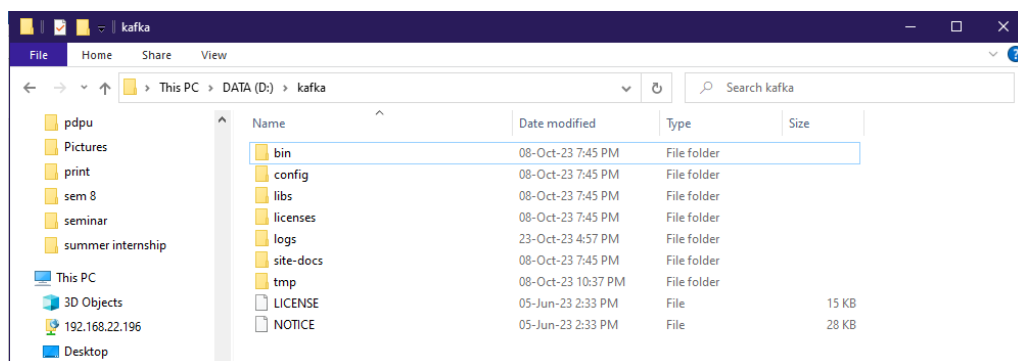
The first step is to download pig form the apache official server , it can be downloaded at following link :

<https://kafka.apache.org/downloads>

In this link download scala 2.13 version under binary downloads.

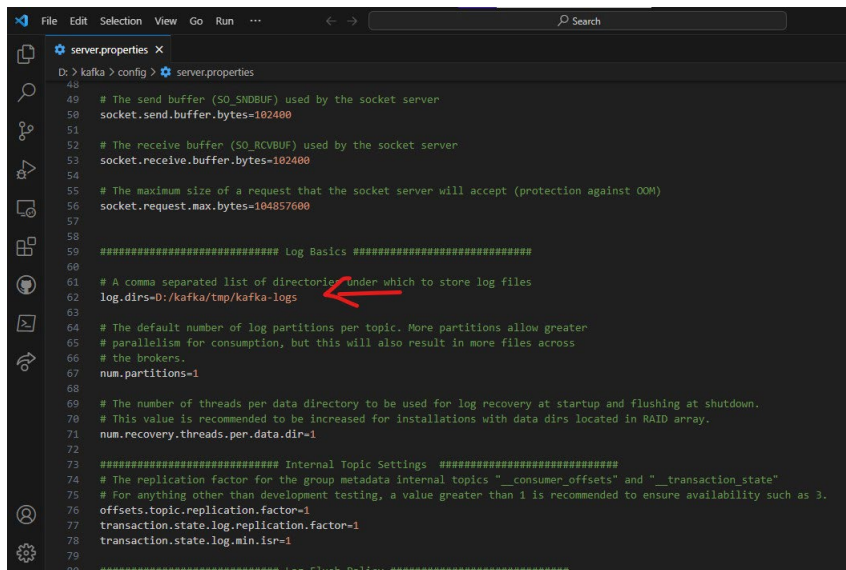


After downloading unzip this folder so it look like this



## ❖ Configure files

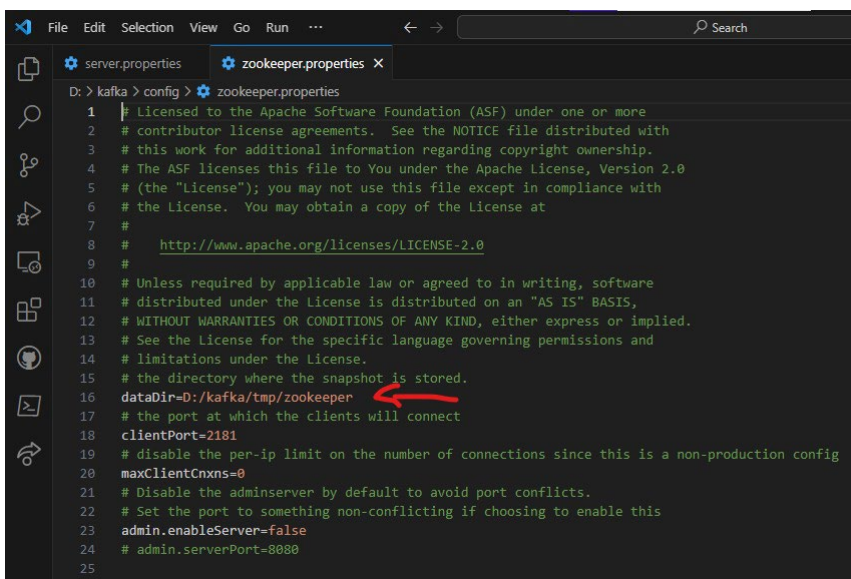
Now go to `kafka>config` and open `server.properties` file , we have to change log directory path to save log on specific location.



```
48
49 # The send buffer (SO_SNDBUF) used by the socket server
50 socket.send.buffer.bytes=102400
51
52 # The receive buffer (SO_RCVBUF) used by the socket server
53 socket.receive.buffer.bytes=102400
54
55 # The maximum size of a request that the socket server will accept (protection against OOM)
56 socket.request.max.bytes=104857600
57
58
59 ##### Log Basics #####
60
61 # A comma separated list of directories under which to store log files
62 log.dirs=D:/kafka/tmp/kafka-logs
63
64 # The default number of log partitions per topic. More partitions allow greater
65 # parallelism for consumption, but this will also result in more files across
66 # the brokers.
67 num.partitions=1
68
69 # The number of threads per data directory to be used for log recovery at startup and flushing at shutdown.
70 # This value is recommended to be increased for installations with data dirs located in RAID array.
71 num.recovery.threads.per.data.dir=1
72
73 ##### Internal Topic Settings #####
74 # The replication factor for the group metadata internal topics "__consumer_offsets" and "__transaction_state"
75 # For anything other than development testing, a value greater than 1 is recommended to ensure availability such as 3.
76 offsets.topic.replication.factor=1
77 transaction.state.log.replication.factor=1
78 transaction.state.log.min.isr=1
79
80 ##### Log Flush Policy #####
```

As shown in above image set `log.dirs` path to your specific location.

Now for the same open `zookeeper.properties` and edit `dataDir` path to specific location as shown below image.



```
1 Licensed to the Apache Software Foundation (ASF) under one or more
2 # contributor license agreements. See the NOTICE file distributed with
3 # this work for additional information regarding copyright ownership.
4 # The ASF licenses this file to You under the Apache License, Version 2.0
5 # (the "License"); you may not use this file except in compliance with
6 # the License. You may obtain a copy of the License at
7 #
8 # http://www.apache.org/licenses/LICENSE-2.0
9 #
10 # Unless required by applicable law or agreed to in writing, software
11 # distributed under the License is distributed on an "AS IS" BASIS,
12 # WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
13 # See the License for the specific language governing permissions and
14 # limitations under the License.
15 # the directory where the snapshot is stored.
16 dataDir=D:/kafka/tmp/zookeeper
17 # the port at which the clients will connect
18 clientPort=2181
19 # disable the per-ip limit on the number of connections since this is a non-production config
20 maxClientCnxns=0
21 # Disable the adminserver by default to avoid port conflicts.
22 # Set the port to something non-conflicting if choosing to enable this
23 admin.enableServer=false
24 # admin.serverPort=8080
25
```

## ❖ Run kafka

To run kafka server first have to start zookeeper server. Now locate to kafka folder where you have bin folder and then open cmd at this place and run following command :

```
.\bin\windows\zookeeper-server-start.bat .\config\zookeeper.properties
```

```
C:\Windows\System32\cmd.exe - .\bin\windows\zookeeper-server-start.bat .\config\zookeeper.properties
[2023-11-16 15:04:47,241] INFO metricsProvider.className is org.apache.zookeeper.metrics.impl.DefaultMetricsProvider (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2023-11-16 15:04:47,241] INFO Starting server (org.apache.zookeeper.server.ZooKeeperServerMain)
[2023-11-16 15:04:47,311] INFO ServerMetrics initialized with provider org.apache.zookeeper.metrics.impl.DefaultMetricsProvider@380fb434 (org.apache.zookeeper.server.ServerMetrics)
[2023-11-16 15:04:47,328] INFO zookeeper.snapshot.trust.empty : false (org.apache.zookeeper.server.persistence.FileTxnSnapLog)
[2023-11-16 15:04:47,406] INFO (org.apache.zookeeper.server.ZooKeeperServer)
[2023-11-16 15:04:47,406] INFO (org.apache.zookeeper.server.ZooKeeperServer)
[2023-11-16 15:04:47,408] INFO (org.apache.zookeeper.server.ZooKeeperServer)
[2023-11-16 15:04:47,409] INFO (org.apache.zookeeper.server.ZooKeeperServer)
[2023-11-16 15:04:47,410] INFO (org.apache.zookeeper.server.ZooKeeperServer)
[2023-11-16 15:04:47,411] INFO (org.apache.zookeeper.server.ZooKeeperServer)
[2023-11-16 15:04:47,412] INFO (org.apache.zookeeper.server.ZooKeeperServer)
[2023-11-16 15:04:47,413] INFO (org.apache.zookeeper.server.ZooKeeperServer)
[2023-11-16 15:04:47,413] INFO (org.apache.zookeeper.server.ZooKeeperServer)
[2023-11-16 15:04:47,414] INFO (org.apache.zookeeper.server.ZooKeeperServer)
[2023-11-16 15:04:47,443] INFO Server environment:zookeeper.version=3.6.4--d65253dcf68e9097c6e95a126463fd5fdeb4521c, built on 12/18/2022 18:10 GMT (org.apache.zookeeper.server.ZooKeeperServer)
[2023-11-16 15:04:47,443] INFO Server environment:host.name=DESKTOP-0007548 (org.apache.zookeeper.server.ZooKeeperServer)
[2023-11-16 15:04:47,445] INFO Server environment:java.version=1.8.0_202 (org.apache.zookeeper.server.ZooKeeperServer)
```

If you see this zookeeper then server started successfully.

Now we have to start kafka server, so open new terminal at same place and run following command :

```
.\bin\windows\kafka-server-start.bat .\config\server.properties
```

Kafka works on producer and consumer model so first we have to create one topic, open new terminal at same place and create a topic by following command :

```
.\bin\windows\kafka-topics.bat --create --topic {topic name} --bootstrap-server localhost:9092
```

Write topic name of your choice don't take space or any special character in topic name it consider only one word of name.

```
Select C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.19045.3570]
(c) Microsoft Corporation. All rights reserved.

D:\kafka>.\bin\windows\kafka-topics.bat --create --topic preExams --bootstrap-server localhost:9092
Created topic preExams.

D:\kafka>
```

After successfully creating a topic, let's start producer by this command, give same topic name that created :

```
.\bin\windows\kafka-console-producer.bat --topic {topic name} --bootstrap-server localhost:9092
```

```
Select C:\Windows\System32\cmd.exe - .\bin\windows\kafka-console-producer.bat --topic preExams --bootstrap-server localhost:9092
Microsoft Windows [Version 10.0.19045.3570]
(c) Microsoft Corporation. All rights reserved.

D:\kafka>.\bin\windows\kafka-topics.bat --create --topic preExams --bootstrap-server localhost:9092
Created topic preExams.

D:\kafka>.\bin\windows\kafka-console-producer.bat --topic preExams --bootstrap-server localhost:9092
>_
```

For the consumer open new terminal and type following command , give same topic name which have given in producer :

`.\bin\windows\kafka-console-consumer.bat --topic {topic name} --from-beginning --bootstrap-server localhost:9092`

```
C:\Windows\System32\cmd.exe - .\bin\windows\kafka-console-consumer.bat --topic preExams --from-beginning --bootstrap-server localhost:9092
Microsoft Windows [Version 10.0.19045.3570]
(c) Microsoft Corporation. All rights reserved.

D:\kafka>.\bin\windows\kafka-console-consumer.bat --topic preExams --from-beginning --bootstrap-server localhost:9092
_
```

Now set producer and consumer side by side , type something in producer and hit enter after some millisecond you can see same thing in consumer terminal.