

Software Requirements

1. SQL Server: <https://www.microsoft.com/en-us/sql-server/sql-server-downloads>

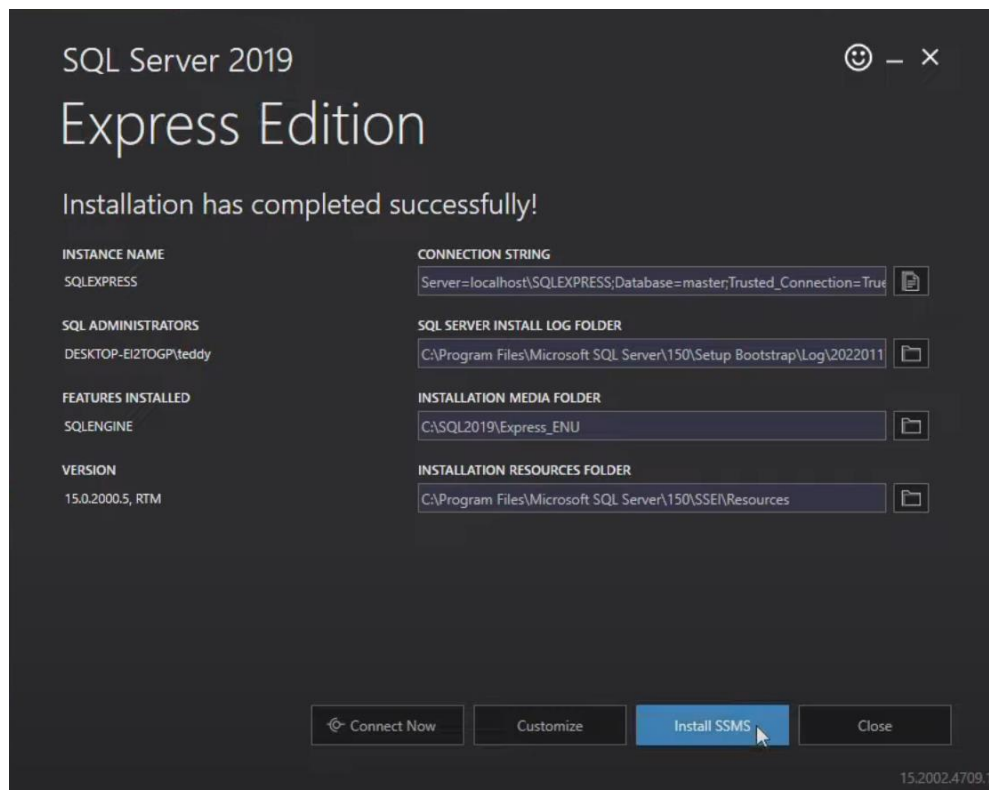


Developer

SQL Server 2022 Developer is a full-featured free edition, licensed for use as a development and test database in a non-production environment.

[Download now](#)

- a. Microsoft SQL Server Management Studio (click Install SSMS when installing SQL Server)



2. Python Version 3.7 or later

- a. Required library: pyodbc → pip install pyodbc

3. Apache Kafka Scala 2.12 or 2.13: <https://kafka.apache.org/downloads>

SUPPORTED RELEASES


3.9.0

- Released November 6, 2024
- [Release Notes](#)
- Docker image: [apache/kafka:3.9.0](#).
- Docker Native image: [apache/kafka-native:3.9.0](#).
- Source download: [kafka-3.9.0-src.tgz](#) ([asc](#), [sha512](#))
- Binary downloads:

- Scala 2.12 - [kafka_2.12-3.9.0.tgz](#) ([asc](#), [sha512](#))
- Scala 2.13 - [kafka_2.13-3.9.0.tgz](#) ([asc](#), [sha512](#))

We build for multiple versions of Scala. This only matters if you are using Scala and you want a version built for the same Scala version you use. Otherwise, any version should work (2.13 is recommended).

4. Apache Zookeeper 3.9.3: <https://zookeeper.apache.org/releases.html>

 Apache ZooKeeper™

Project • Documentation • Developers • ASF •

Apache ZooKeeper™ Releases

The Apache ZooKeeper system for distributed coordination is a high-performance service for building distributed applications.

- [Release strategy](#)
- [Download](#)
- [Verifying Hashes and Signatures](#)
- [News](#)

Release strategy

The Apache ZooKeeper community supports two release branches at a time: **stable** and **current**. The **stable** version of ZooKeeper is 3.8.x and the **current** version is 3.9.x. Once a new minor version is released, the **stable** version is expected to be decommissioned soon and in approximately half a year will be announced as End-of-Life. During the half year grace period only security and critical fixes are expected to be released for the version. After EoL is announced no further patches are provided by the community. All ZooKeeper releases will remain accessible from the official Apache Archives.

Download

Apache ZooKeeper 3.9.3 is our current release, and 3.8.4 our latest stable release.

Apache ZooKeeper 3.9.3

[Apache ZooKeeper 3.9.3\(asc, sha512\)](#)
[Apache ZooKeeper 3.9.3 Source Release\(asc, sha512\)](#)

Apache ZooKeeper 3.8.4 (latest stable release)

[Apache ZooKeeper 3.8.4\(asc, sha512\)](#)
[Apache ZooKeeper 3.8.4 Source Release\(asc, sha512\)](#)

Apache ZooKeeper 3.7.2 (3.7 is EoL since 2nd of February, 2024)

[Apache ZooKeeper 3.7.2\(asc, sha512\)](#)
[Apache ZooKeeper 3.7.2 Source Release\(asc, sha512\)](#)

Older releases are available in the archive.

Setup Kafka

Instructions found here: <https://www.geeksforgeeks.org/how-to-install-and-run-apache-kafka-on-windows/>

1. Extract Apache Kafka your desired directory.
2. Copy the path of the Kafka folder. Now go to *config* inside kafka folder and open *zookeeper.properties* file. Copy the path against the field *dataDir* and add */zookeeper-data* to the path.

```
zookeeper.properties X
C: > Kafka > kafka_2.12-3.9.0 > config > zookeeper.properties
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=C:\Kafka\kafka 2.12-3.9.0\zookeeper-data
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
```

- Now in the same folder *config* open *server.properties* and scroll down to *log.dirs* and paste the path. To the path add */kafka-logs*

```
server.properties
C:\Kafka> kafka-2.12-3.0.0\config> type server.properties
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
16 #
17 # This configuration file is intended for use in zk-based mode, where Apache Zookeeper is required.
18 # See kafka.server.KafkaConfig for additional details and defaults
19 #
20
21 ##### Server Basics #####
22
23 # The id of the broker. This must be set to a unique integer for each broker.
24 broker.id=0
25
26 ##### Socket Server Settings #####
27
28 # The address the socket server listens on. If not configured, the host name will be equal to the value of
29 # java.net.InetAddress.getCanonicalHostName(), with PLAINTEXT listener name, and port 9092.
30 #
31 # listeners = listener_name://host_name:port
32 #
33 # EXAMPLE:
34 # listeners = PLAINTEXT://your.host.name:9092
35 # listeners=PLAINTEXT://:9092
36
37 # listener name, hostname and port the broker will advertise to clients.
38 # If not set, it uses the value for "listeners".
39 #advertised.listeners=PLAINTEXT://your.host.name:9092
40
41 # Maps listener names to security protocols, the default is for them to be the same. See the config documentation for more details
42 #listener.security.protocol.map=PLAINTEXT:PLAINTEXT,SSL:SSL,SASL_PLAINTEXT:SASL_PLAINTEXT,SASL_SSI:SASL_SSI
43
44 # The number of threads that the server uses for receiving requests from the network and sending responses to the network
45 num.network.threads=3
46
47 # The number of threads that the server uses for processing requests, which may include disk I/O
48 num.io.threads=8
49
50 # The send buffer (SO_SNDBUF) used by the socket server
51 socket.send.buffer.bytes=102400
52
53 # The receive buffer (SO_RCVBUF) used by the socket server
54 socket.receive.buffer.bytes=102400
55
56 # The maximum size of a request that the socket server will accept (protection against OOM)
57 socket.request.max.bytes=104857600
58
59 ##### Log Basics #####
60
61 # A comma separated list of directories under which to store log files
62 log.dirs=C:\Kafka\kafka-2.12-3.0.0\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
```

- Now open command prompt and change the directory to the kafka folder. First start zookeeper using the command given below:

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

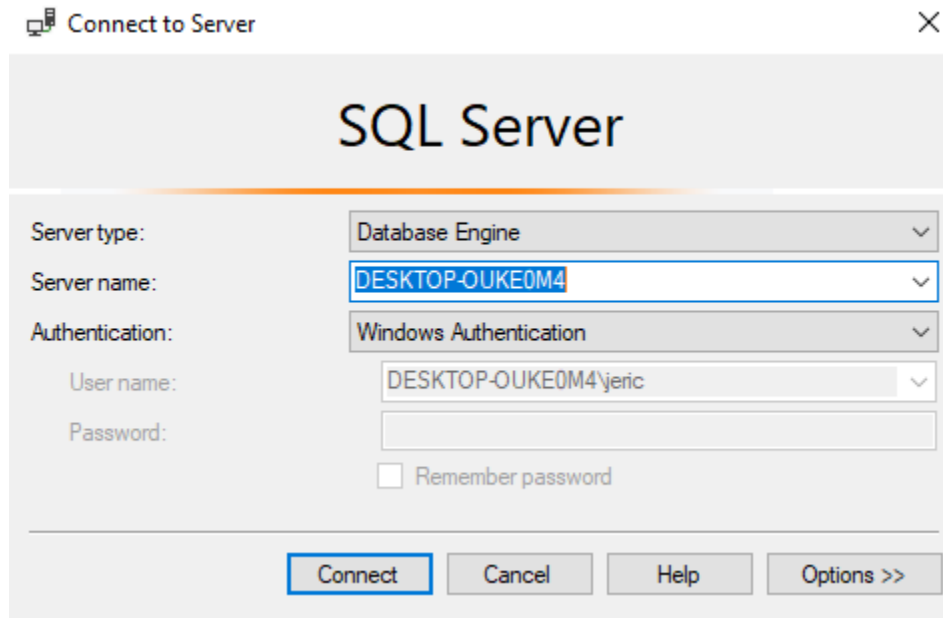
```
C:\Kafka\kafka-2.12-3.0.0>.\bin\windows\zookeeper-server-start.bat .\config\zookeeper.properties
2024-11-13 16:15:47.149 INFO Reading configuration from: .\config\zookeeper.properties (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
2024-11-13 16:15:47.154 WARN C:\Kafka\kafka-2.12-3.0.0\zookeeper-data is relative. Prepend .\ to indicate that you're sure! (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
2024-11-13 16:15:47.156 INFO clientPortAddress is 0.0.0.0:2181 (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
2024-11-13 16:15:47.156 INFO secureClientPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
2024-11-13 16:15:47.156 INFO observerMasterPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
2024-11-13 16:15:47.156 INFO metricsProvider.className is org.apache.zookeeper.metrics.impl.DefaultMetricsProvider (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
2024-11-13 16:15:47.157 INFO autopurge.snapRetainCount set to 3 (org.apache.zookeeper.server.DataDirCleanupManager)
2024-11-13 16:15:47.158 INFO autopurge.purgeInterval set to 0 (org.apache.zookeeper.server.DataDirCleanupManager)
2024-11-13 16:15:47.158 INFO Purge task is not scheduled. (org.apache.zookeeper.server.DataDirCleanupManager)
2024-11-13 16:15:47.158 WARN Either no config or no quorum defined in config, running in standalone mode (org.apache.zookeeper.server.quorum.QuorumPeerMain)
2024-11-13 16:15:47.159 INFO Log4j 1.2 Jmx support not found; Jmx disabled. (org.apache.zookeeper.JmxManagedUtil)
2024-11-13 16:15:47.161 INFO Reading configuration from: .\config\zookeeper.properties (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
2024-11-13 16:15:47.162 WARN C:\Kafka\kafka-2.12-3.0.0\zookeeper-data is relative. Prepend .\ to indicate that you're sure! (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
2024-11-13 16:15:47.162 INFO clientPortAddress is 0.0.0.0:2181 (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
2024-11-13 16:15:47.162 INFO secureClientPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
2024-11-13 16:15:47.162 INFO observerMasterPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
2024-11-13 16:15:47.163 INFO metricsProvider.className is org.apache.zookeeper.metrics.impl.DefaultMetricsProvider (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
2024-11-13 16:15:47.163 INFO Starting server (org.apache.zookeeper.server.ZooKeeperServerMain)
2024-11-13 16:15:47.172 INFO ServerMetrics initialized with provider org.apache.zookeeper.metrics.impl.DefaultMetricsProvider@6c19a4 (org.apache.zookeeper.server.ServerMetrics)
2024-11-13 16:15:47.179 INFO ACL digest algorithm is: SHA1 (org.apache.zookeeper.server.auth.DigestAuthenticationProvider)
2024-11-13 16:15:47.179 INFO zookeeper.digestAuthenticationProvider.enabled = true (org.apache.zookeeper.server.auth.DigestAuthenticationProvider)
2024-11-13 16:15:47.183 INFO zookeeper.snapshot.trust.empty = false (org.apache.zookeeper.server.persistence.FileTxnSnapLog)
2024-11-13 16:15:47.194 INFO (org.apache.zookeeper.server.ZooKeeperServer)
2024-11-13 16:15:47.195 INFO (org.apache.zookeeper.server.ZooKeeperServer)
2024-11-13 16:15:47.196 INFO (org.apache.zookeeper.server.ZooKeeperServer)
2024-11-13 16:15:47.196 INFO (org.apache.zookeeper.server.ZooKeeperServer)
2024-11-13 16:15:47.197 INFO (org.apache.zookeeper.server.ZooKeeperServer)
2024-11-13 16:15:47.197 INFO (org.apache.zookeeper.server.ZooKeeperServer)
2024-11-13 16:15:47.197 INFO (org.apache.zookeeper.server.ZooKeeperServer)
2024-11-13 16:15:47.198 INFO (org.apache.zookeeper.server.ZooKeeperServer)
2024-11-13 16:15:47.200 INFO Server environment:zookeeper.version=3.4.4-916c2a707a166dd8f6593f3add6c36cc436c, built on 2024-02-12 22:16 UTC (org.apache.zookeeper.server.ZooKeeperServer)
2024-11-13 16:15:47.200 INFO Server environment:host.name=DESKTOP-DASHOME ashome.net (org.apache.zookeeper.server.ZooKeeperServer)
2024-11-13 16:15:47.201 INFO Server environment:java.version=17.0.9 (org.apache.zookeeper.server.ZooKeeperServer)
2024-11-13 16:15:47.201 INFO Server environment:java.vendor=Oracle Corporation (org.apache.zookeeper.server.ZooKeeperServer)
```

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

[illegible]

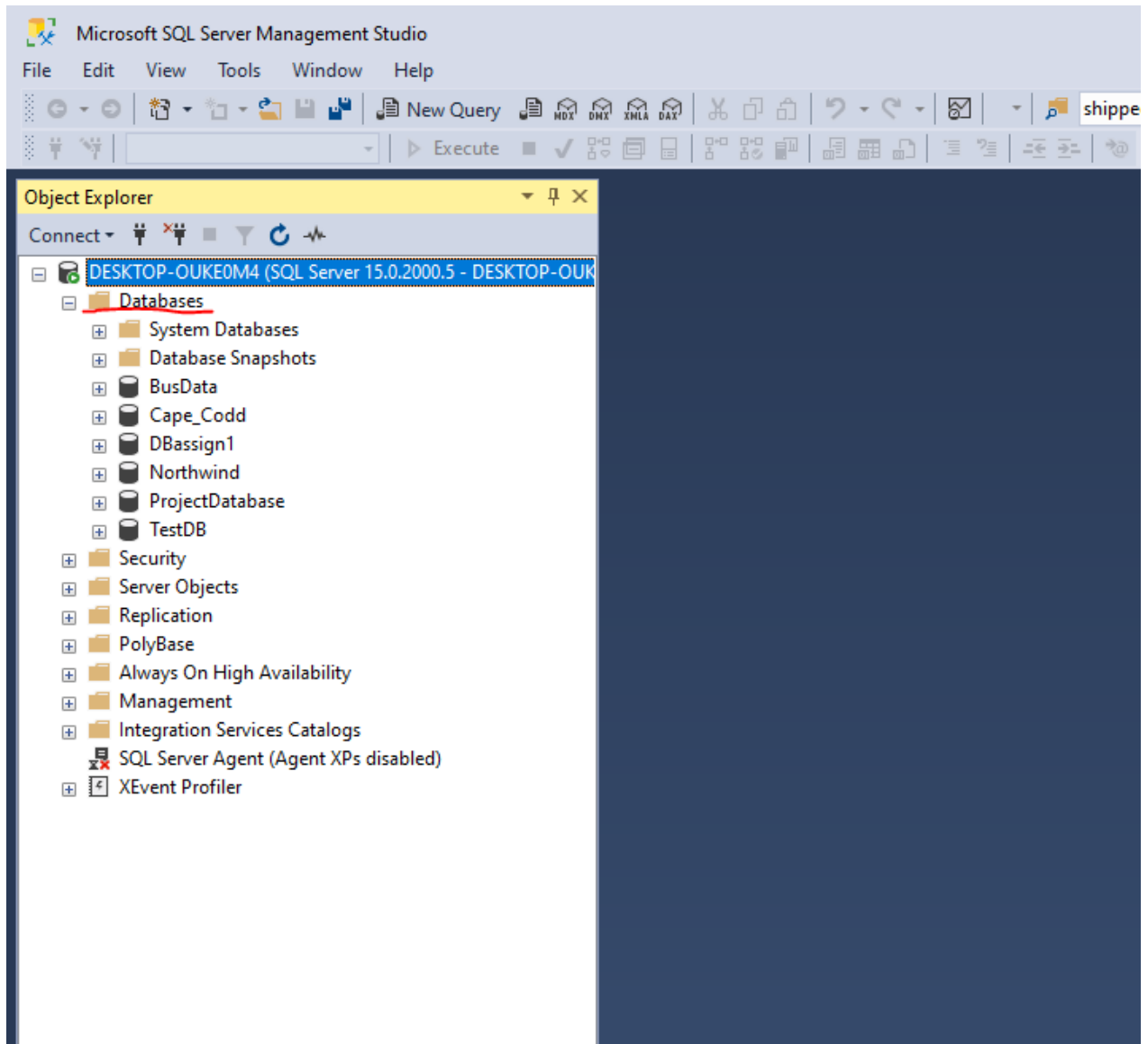
# Setup Microsoft SQL Server Management Studio

1. Launch Microsoft SQL Server Management Studio



2. Take note of the Server name and connect.

3. Create a new database, right click on 'Databases' and click create new 'Database'...



#### 4. Name your database and click OK

**New Database**

Script ? Help

**Select a page**

- General
- Options
- Filegroups

**Connection**

Server:  
DESKTOP-OUKE0M4

Connection:  
DESKTOP-OUKE0M4\jeric

[View connection properties](#)

**Progress**

Ready

Database name:

Owner:

☒ Use full-text indexing

Database files:

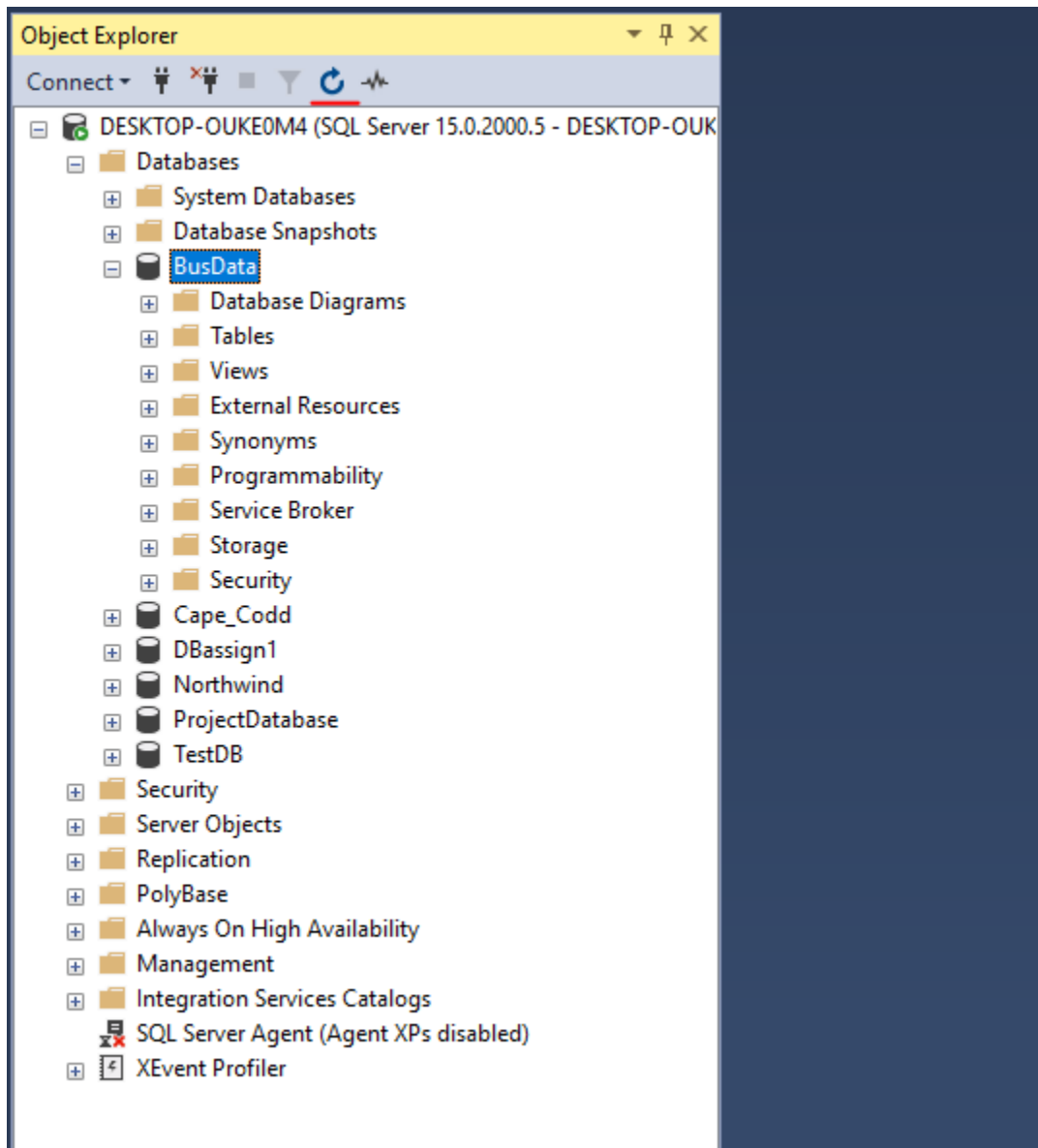
| Logical Name | File Type | Filegroup      | Initial Size (MB) | Autogrowth / Maxsize | Path   |
|--------------|-----------|----------------|-------------------|----------------------|--------|
| ROWS...      | ROWS...   | PRIMARY        | 8                 | By 64 MB, Unlimited  | C:\... |
| _log         | LOG       | Not Applicable | 8                 | By 64 MB, Unlimited  | C:\... |

Add Remove

OK Cancel

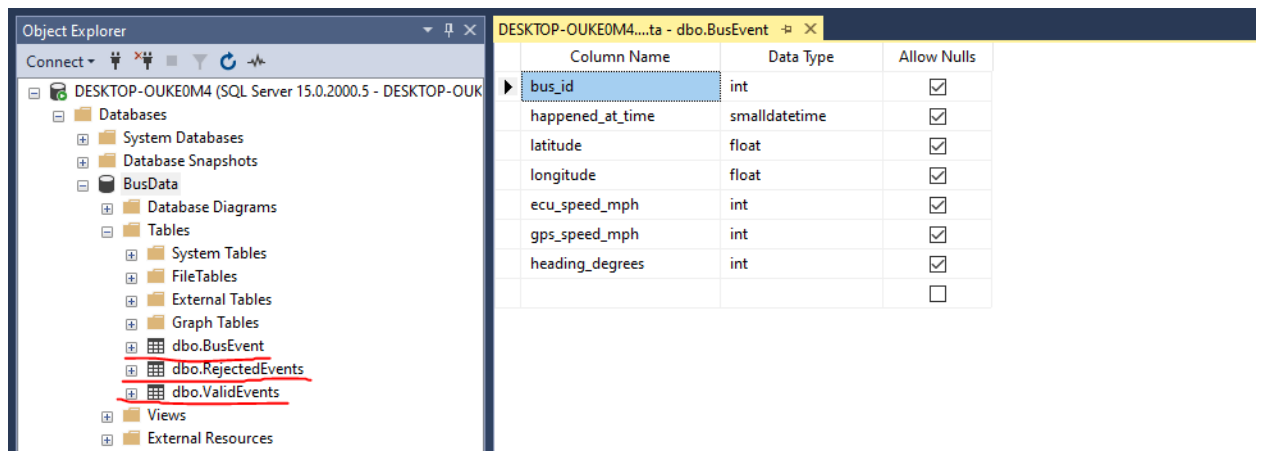


5. Refresh to see your database.



6. Expand your database you created by double clicking it, or by clicking the + next to your database
7. Create three new tables by right clicking 'Tables', clicking 'New' and then Table...
8. Each table will have the same column names and data types.
  - a. bus\_id: int
  - b. happened\_at\_time: smalldatetime
  - c. latitude: float
  - d. longitude: float
  - e. ecu\_speed\_mph: int
  - f. gps\_speed\_mph: int
  - g. heading\_degrees: int

|   | Column Name      | Data Type     | Allow Nulls                         |
|---|------------------|---------------|-------------------------------------|
| ► | bus_id           | int           | <input checked="" type="checkbox"/> |
|   | happened_at_time | smalldatetime | <input checked="" type="checkbox"/> |
|   | latitude         | float         | <input checked="" type="checkbox"/> |
|   | longitude        | float         | <input checked="" type="checkbox"/> |
|   | ecu_speed_mph    | int           | <input checked="" type="checkbox"/> |
|   | gps_speed_mph    | int           | <input checked="" type="checkbox"/> |
|   | heading_degrees  | int           | <input checked="" type="checkbox"/> |
|   |                  |               | <input type="checkbox"/>            |



9. Setup for Microsoft SQL Server Management Studio is complete, take note of what the Server Name is, what your database name is, and the name of your three tables.

# Running the Source Code

1. Copy and Paste the Source Code from 'BusDataGenerationAndInsertion.py' save and name your file.
2. Adjust the names of the server and the database to match your own.

```
1 import pyodbc
2 import random
3 import time
4 from datetime import datetime, timedelta
5 from kafka import KafkaProducer
6 import json
7
8 # Kafka connection details
9 KAFKA_TOPIC = 'bus_data'
10 KAFKA_SERVER = 'localhost:9092' # Update this to your Kafka server if different
11
12 # SQL Server connection details
13 server = 'DESKTOP-OUKEOM4'
14 database = 'BusData'
15 driver = '{ODBC Driver 17 for SQL Server}'
16
17 # Initialize Kafka Producer
18 producer = KafkaProducer(
19 bootstrap_servers=KAFKA_SERVER,
20 value_serializer=lambda v: json.dumps(v, default=str).encode('utf-8')
21)
22
23 # Connect to SQL Server
24 connection_string = f'DRIVER={driver};SERVER={server};DATABASE={database};Trusted_Connection=yes;'
25 conn = pyodbc.connect(connection_string)
26 cursor = conn.cursor()
27 print("Connected to SQL Server and Kafka.")
28
29
30 # Generate and insert random data
31 usage
32 def generate_random_data():
33 bus_id = random.randint(a=1, b=100)
34 happened_at_time = datetime.now() - timedelta(minutes=random.randint(a=0, b=1000))
35 latitude = round(random.uniform(-90, b=90), 6)
36 longitude = round(random.uniform(-180, b=180), 6)
37 ecu_speed_mph = random.randint(a=0, b=120)
38 gps_speed_mph = random.randint(a=0, b=120)
39 heading_degrees = random.randint(a=0, b=360)
40
41 row_data = {
42 "bus_id": bus_id,
43 "happened_at_time": happened_at_time,
44 "latitude": latitude,
45 "longitude": longitude,
46 "ecu_speed_mph": ecu_speed_mph,
47 "gps_speed_mph": gps_speed_mph,
48 "heading_degrees": heading_degrees
49 }
```

3. Adjust the names of the databases to match your own.

```

50 # Determine if data is valid or rejected based on speed criteria
51 if ecu_speed_mph > 70 or gps_speed_mph > 70:
52 cursor.execute(sql: '''
53 INSERT INTO dbo.RejectedEvents (bus_id, happened_at_time, latitude, longitude, ecu_speed_mph, gps_speed_mph, heading_degrees)
54 VALUES (?, ?, ?, ?, ?, ?, ?)
55 ''', *params: (bus_id, happened_at_time, latitude, longitude, ecu_speed_mph, gps_speed_mph, heading_degrees))
56 print("An invalid record was inserted into RejectedEvents.")
57 else:
58 cursor.execute(sql: '''
59 INSERT INTO dbo.BusEvent (bus_id, happened_at_time, latitude, longitude, ecu_speed_mph, gps_speed_mph, heading_degrees)
60 VALUES (?, ?, ?, ?, ?, ?, ?)
61 ''', *params: (bus_id, happened_at_time, latitude, longitude, ecu_speed_mph, gps_speed_mph, heading_degrees))
62
63 cursor.execute(sql: '''
64 INSERT INTO dbo.ValidEvents (bus_id, happened_at_time, latitude, longitude, ecu_speed_mph, gps_speed_mph, heading_degrees)
65 VALUES (?, ?, ?, ?, ?, ?, ?)
66 ''', *params: (bus_id, happened_at_time, latitude, longitude, ecu_speed_mph, gps_speed_mph, heading_degrees))
67 print("A valid record was inserted into BusEvent and ValidEvents.")
68

```

4. The program will now run properly when zookeeper and a Kafka server are running.