Assignment 4 ⇒

```
Last login: Sat Nov  8 16:19:27 on ttys001
bhavychawla@Bhavys-MacBook-Pro ~ % open -a TextEdit /opt/homebrew/opt/kafka/libexec/config/server.properties

bhavychawla@Bhavys-MacBook-Pro ~ % kafka-topics.sh --create \
  --topic wiki-vote \
  --bootstrap-server localhost:9092 \
  --partitions 1 \
  --replication-factor 1

Created topic wiki-vote.
bhavychawla@Bhavys-MacBook-Pro ~ % kafka-topics.sh --list --bootstrap-server localhost:9092

wiki-vote
bhavychawla@Bhavys-MacBook-Pro ~ % curl -O https://snap.stanford.edu/data/wiki-Vote.txt.gz
gunzip wiki-Vote.txt.gz

  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100  283k  100  283k    0     0  62467      0  0:00:04  0:00:04 --:--:-- 67947
bhavychawla@Bhavys-MacBook-Pro ~ % nano producer.py
```

Configure apache Kafka by first editing server.properties file, set local host: 9092 as server with prt 9092 for Kafka server daemon to start.

Then create a topic wiki-vote as channel to hold streams of data, ie named Stream of Records!

As well as download the dataset using curl.

```
bhavychawla@Bhavys-MacBook-Pro ~ % kafka-server-start.sh /opt/homebrew/opt/kafka/libexec/config/server.properties

[2025-11-08 19:55:58,603] INFO Registered `kafka:type=kafka.Log4jController` MBean (kafka.utils.Log4jControllerRegistration$)
[2025-11-08 19:55:58,868] INFO Registered signal handlers for TERM, INT, HUP (org.apache.kafka.common.utils.LoggingSignalHandler)
[2025-11-08 19:55:58,869] INFO [ControllerServer id=1] Starting controller (kafka.server.ControllerServer)
[2025-11-08 19:55:59,082] INFO Updated connection-accept-rate max connection creation rate to 2147483647 (kafka.network.ConnectionQuotas)
[2025-11-08 19:55:59,096] INFO [SocketServer listenerType=CONTROLLER, nodeId=1] Created data-plane acceptor and processors for endpoint : ListenerName(CONTROLLER) (kafka.network.SocketServer)
[2025-11-08 19:55:59,098] INFO authorizerStart completed for endpoint CONTROLLER. Endpoint is now READY. (org.apache.kafka.server.network.EndpointReadyFutures)
[2025-11-08 19:55:59,099] INFO [SharedServer id=1] Starting SharedServer (kafka.server.SharedServer)
[2025-11-08 19:55:59,121] INFO [LogLoader partition=__cluster_metadata-0, dir=/opt/homebrew/var/lib/kraft-combined-logs] Recovering unflushed segment 0. 0 recovered for __cluster_metadata-0. (org.apache.kafka.storage.internals.log.LogLoader)
[2025-11-08 19:55:59,122] INFO [LogLoader partition=__cluster_metadata-0, dir=/opt/homebrew/var/lib/kraft-combined-logs] Loading producer state till offset 0 (org.apache.kafka.storage.internals.log.UnifiedLog)
[2025-11-08 19:55:59,122] INFO [LogLoader partition=__cluster_metadata-0, dir=/opt/homebrew/var/lib/kraft-combined-logs] Reloading from producer snapshot and rebuilding producer state from offset 0 (org.apache.kafka.storage.internals.log.UnifiedLog)
[2025-11-08 19:55:59,123] INFO Deleted producer state snapshot /opt/homebrew/var/lib/kraft-combined-logs/__cluster_metadata-0/00000000000000013609.snapshot (org.apache.kafka.storage.internals.log.SnapshotFile)
[2025-11-08 19:55:59,123] INFO Deleted producer state snapshot /opt/homebrew/var/lib/kraft-combined-logs/__cluster_metadata-0/00000000000000013862.snapshot (org.apache.kafka.storage.internals.log.SnapshotFile)
[2025-11-08 19:55:59,123] INFO [LogLoader partition=__cluster_metadata-0, dir=/opt/homebrew/var/lib/kraft-combined-logs] Producer state recovery took 1ms for snapshot load and 0ms for segment recovery from offset 0 (org.apache.kafka.storage.internals.log.UnifiedLog)
[2025-11-08 19:55:59,162] INFO [ProducerStateManager partition=__cluster_metadata-0] Wrote producer snapshot at offset 13862 with 0 producer ids in 7 ms. (org.apache.kafka.storage.internals.log.ProducerStateManager)
[2025-11-08 19:55:59,164] INFO [LogLoader partition=__cluster_metadata-0, dir=/opt/homebrew/var/lib/kraft-combined-logs] Loading producer state till offset 13862 (org.apache.kafka.storage.internals.log.UnifiedLog)
[2025-11-08 19:55:59,164] INFO [LogLoader partition=__cluster_metadata-0, dir=/opt/homebrew/var/lib/kraft-combined-logs] Reloading from producer snapshot and rebuilding producer state from offset 13862 (org.apache.kafka.storage.internals.log.UnifiedLog)
[2025-11-08 19:55:59,165] INFO [ProducerStateManager partition=__cluster_metadata-0] Loading producer state from snapshot file 'SnapshotFile(offset=13862, file=/opt/homebrew/var/lib/kraft-combined-logs/__cluster_metadata-0/00000000000000013862.snapshot)' (org.apache.kafka.storage.internals.log.ProducerStateManager)
[2025-11-08 19:55:59,165] INFO [LogLoader partition=__cluster_metadata-0, dir=/opt/homebrew/var/lib/kraft-combined-logs] Producer state recovery took 1ms for snapshot load and 0ms for segment recovery from offset 13862 (org.apache.kafka.storage.internals.log.UnifiedLog)
[2025-11-08 19:55:59,179] INFO Initialized snapshots with IDs SortedSet(OffsetAndEpoch[offset=7161, epoch=1]) from /opt/homebrew/var/lib/kraft-combined-logs/__cluster_metadata-0 (kafka.raft.KafkaMetadataLog$)
[2025-11-08 19:55:59,184] INFO [raft-expiration-reaper]: Starting (org.apache.kafka.raft.TimingWheelExpirationService$ExpiredOperationReaper)
[2025-11-08 19:55:59,191] INFO [RaftManager id=1] Starting request manager with bootstrap servers: [localhost:9093 (id: -2 rack: null isFenced: false)] (org.apache.kafka.raft.KafkaRaftClient)
[2025-11-08 19:55:59,193] INFO [RaftManager id=1] Reading KRaft snapshot and log as part of the initialization (org.apache.kafka.raft.KafkaRaftClient)
[2025-11-08 19:55:59,194] INFO [RaftManager id=1] Loading snapshot (OffsetAndEpoch[offset=7161, epoch=1]) since log start offset (0) is greater than the internal listener's next offset (-1) (org.apache.kafka.raft.internals.KRaftControlRecordStateMachine)
[2025-11-08 19:55:59,209] INFO [RaftManager id=1] Starting voters are VoterSet(voters={1=VoterNode(voterKey=ReplicaKey(id=1, directoryId=<undefined>), listeners=Endpoints(endpoints={ListenerName(CONTROLLER)=localhost/127.0.0.1:9093}), supportedKRaftVersion=SupportedVersionRange[min_version:0, max_version:0])}) (org.apache.kafka.raft.KafkaRaftClient)
[2025-11-08 19:55:59,214] INFO [RaftManager id=1] Attempting durable transition to ResignedState(localId=1, epoch=5, voters=[1], electionTimeoutMs=1456, unackedVoters=[], preferredSuccessors=[]) from null (org.apache.kafka.raft.QuorumState)
[2025-11-08 19:55:59,224] INFO [RaftManager id=1] Completed transition to ResignedState(localId=1, epoch=5, voters=[1], electionTimeoutMs=1456, unackedVoters=[], preferredSuccessors=[]) from null (org.apache.kafka.raft.QuorumState)
[2025-11-08 19:55:59,226] INFO [RaftManager id=1] Completed transition to ProspectiveState(epoch=5, leaderId=OptionalInt[1], votedKey=Optional.empty, epochElection=EpochElection(voterStates={1=VoterState(replicaKey=ReplicaKey(id=1, directoryId=<undefined>), state=GRANTED)}), electionTimeoutMs=1337, highWatermark=Optional.empty) from ResignedState(localId=1, epoch=5, voters=[1], electionTimeoutMs=1456, unackedVoters=[], preferredSuccessors=[]) (org.apache.kafka.raft.QuorumState)
[2025-11-08 19:55:59,226] INFO [RaftManager id=1] Attempting durable transition to CandidateState(localId=1, localDirectoryId=_QaHqb2s1M_RbZEtzGO0KQ, epoch=6, epochElection=EpochElection(voterStates={1=VoterState(replicaKey=ReplicaKey(id=1, directoryId=<undefined>), state=GRANTED)}), highWatermark=Optional.empty, electionTimeoutMs=1175) from ProspectiveState(epoch=5, leaderId=OptionalInt[1], votedKey=Optional.empty, epochElection=EpochElection(voterStates={1=VoterState(replicaKey=ReplicaKey(id=1, directoryId=<undefined>), state=GRANTED)}), electionTimeoutMs=1337, highWatermark=Optional.empty) (org.apache.kafka.raft.QuorumState)
[2025-11-08 19:55:59,230] INFO [RaftManager id=1] Completed transition to CandidateState(localId=1, localDirectoryId=_QaHqb2s1M_RbZEtzGO0KQ, epoch=6, epochElection=EpochElection(voterStates={1=VoterState(replicaKey=ReplicaKey(id=1, directoryId=<undefined>), state=GRANTED)}), highWatermark=Optional.empty, electionTimeoutMs=1337, highWatermark=Optional.empty) (org.apache.kafka.raft.QuorumState)
[2025-11-08 19:55:59,231] INFO [RaftManager id=1] Attempting durable transition to Leader(localVoterNode=VoterNode(voterKey=ReplicaKey(id=1, directoryId=_QaHqb2s1M_RbZEtzGO0KQ), listeners=Endpoints(endpoints={ListenerName(CONTROLLER)=localhost/<unresolved>:9093}), supportedKRaftVersion=SupportedVersionRange[min_version:0, max_version:1]), epoch=6, epochStartOffset=13862, highWatermark=Optional.empty, voterStates={1=ReplicaState(replicaKey=ReplicaKey(id=1, directoryId=<undefined>), endOffset=Optional.empty, lastFetchTimestamp=-1, lastCaughtUpTimestamp=-1, hasAcknowledgedLeader=true)}) from CandidateState(localId=1, localDirectoryId=_QaHqb2s1M_RbZEtzGO0KQ, epoch=6, epochElection=EpochElection(voterStates={1=VoterState(replicaKey=ReplicaKey(id=1, directoryId=<undefined>), state=GRANTED)}), highWatermark=Optional.empty, electionTimeoutMs=1175) (org.apache.kafka.raft.QuorumState)
[2025-11-08 19:55:59,236] INFO [RaftManager id=1] Completed transition to Leader(localVoterNode=VoterNode(voterKey=ReplicaKey(id=1, directoryId=_QaHqb2s1M_RbZEtzGO0KQ), listeners=Endpoints(endpoints={ListenerName(CONTROLLER)=localhost/<unresolved>:9093}), supportedKRaftVersion=SupportedVersionRange[min_version:0, max_version:1]), epoch=6, epochStartOffset=13862, highWatermark=Optional.empty, voterStates={1=ReplicaState(replicaKey=ReplicaKey(id=1, directoryId=<undefined>), endOffset=Optional.empty, lastFetchTimestamp=-1, lastCaughtUpTimestamp=-1, hasAcknowledgedLeader=true)}) from CandidateState(localId=1, localDirectoryId=_QaHqb2s1M_RbZEtzGO0KQ, epoch=6, epochElection=EpochElection(voterStates={1=VoterState(replicaKey=ReplicaKey(id=1, directoryId=<undefined>), state=GRANTED)}), highWatermark=Optional.empty, electionTimeoutMs=1175) (org.apache.kafka.raft.QuorumState)
[2025-11-08 19:55:59,242] INFO [kafka-1-raft-outbound-request-thread]: Starting (org.apache.kafka.raft.KafkaNetworkChannel$SendThread)
[2025-11-08 19:55:59,242] INFO [kafka-1-raft-io-thread]: Starting (org.apache.kafka.raft.KafkaRaftClientDriver)
[2025-11-08 19:55:59,247] INFO [RaftManager id=1] High watermark set to LogOffsetMetadata(offset=13863, metadata=Optional[(segmentBaseOffset=0,relativePositionInSegment=991111)]) for the first time for epoch 6 based on indexOfHw 0 and voters [ReplicaState(replicaKey=ReplicaKey(id=1, directoryId=<undefined>), endOffset=Optional[LogOffsetMetadata(offset=13863, metadata=Optional[(segmentBaseOffset=0,relativePositionInSegment=991111)])], lastFetchTimestamp=-1, lastCaughtUpTimestamp=-1, hasAcknowledgedLeader=true)] (org.apache.kafka.raft.LeaderState)
[2025-11-08 19:55:59,248] INFO [MetadataLoader id=1] initializeNewPublishers: The loader is still catching up because we have loaded up to offset -1, but the high water mark is 13863 (org.apache.kafka.image.loader.MetadataLoader)
```

```
Password:
Sorry, try again.
Password:
bhavychawla@Bhavys-MacBook-Pro ~ % mkdir -p /opt/homebrew/var/lib/kraft-combined-logs

bhavychawla@Bhavys-MacBook-Pro ~ % kafka-storage.sh format --config /opt/homebrew/opt/kafka/libexec/config/server.properties --cluster-id "$(kafka-storage.sh random-uuid)" --ignore-formatted

Formatting metadata directory /opt/homebrew/var/lib/kraft-combined-logs with metadata.version 4.1-IV1.
bhavychawla@Bhavys-MacBook-Pro ~ % kafka-server-start.sh /opt/homebrew/opt/kafka/libexec/config/server.properties

[2025-11-08 17:50:20,475] INFO Registered `kafka:type=kafka.Log4jController` MBean (kafka.utils.Log4jControllerRegistration$)
[2025-11-08 17:50:20,616] INFO Registered signal handlers for TERM, INT, HUP (org.apache.kafka.common.utils.LoggingSignalHandler)
[2025-11-08 17:50:20,616] INFO [ControllerServer id=1] Starting controller (kafka.server.ControllerServer)
[2025-11-08 17:50:20,735] INFO Updated connection-accept-rate max connection creation rate to 2147483647 (kafka.network.ConnectionQuotas)
[2025-11-08 17:50:20,748] INFO [SocketServer listenerType=CONTROLLER, nodeId=1] Created data-plane acceptor and processors for endpoint : ListenerName(CONTROLLER) (kafka.network.SocketServer)
[2025-11-08 17:50:20,751] INFO authorizerStart completed for endpoint CONTROLLER. Endpoint is now READY. (org.apache.kafka.server.network.EndpointReadyFutures)
[2025-11-08 17:50:20,751] INFO [SharedServer id=1] Starting SharedServer (kafka.server.SharedServer)
[2025-11-08 17:50:20,772] INFO [LogLoader partition=__cluster_metadata-0, dir=/opt/homebrew/var/lib/kraft-combined-logs] Loading producer state till offset 0 (org.apache.kafka.storage.internals.log.UnifiedLog)
[2025-11-08 17:50:20,772] INFO [LogLoader partition=__cluster_metadata-0, dir=/opt/homebrew/var/lib/kraft-combined-logs] Reloading from producer snapshot and rebuilding producer state from offset 0 (org.apache.kafka.storage.internals.log.UnifiedLog)
[2025-11-08 17:50:20,772] INFO [LogLoader partition=__cluster_metadata-0, dir=/opt/homebrew/var/lib/kraft-combined-logs] Producer state recovery took 0ms for snapshot load and 0ms for segment recovery from offset 0 (org.apache.kafka.storage.internals.log.UnifiedLog)
[2025-11-08 17:50:20,781] INFO Initialized snapshots with IDs SortedSet() from /opt/homebrew/var/lib/kraft-combined-logs/__cluster_metadata-0 (kafka.raft.KafkaMetadataLog$)
[2025-11-08 17:50:20,785] INFO [raft-expiration-reaper]: Starting (org.apache.kafka.raft.TimingWheelExpirationService$ExpiredOperationReaper)
[2025-11-08 17:50:20,798] INFO [RaftManager id=1] Starting request manager with bootstrap servers: [localhost:9093 (id: -2 rack: null isFenced: false)] (org.apache.kafka.raft.KafkaRaftClient)
[2025-11-08 17:50:20,792] INFO [RaftManager id=1] Reading KRaft snapshot and log as part of the initialization (org.apache.kafka.raft.KafkaRaftClient)
[2025-11-08 17:50:20,793] INFO [RaftManager id=1] Starting voters are VoterSet(voters={1=VoterNode(voterKey=ReplicaKey(id=1, directoryId=<undefined>), listeners=Endpoints(endpoints={ListenerName(CONTROLLER)=localhost/127.0.0.1:9093}), supportedKRaftVersion=SupportedVersionRange[min_version:0, max_version:0])}) (org.apache.kafka.raft.KafkaRaftClient)
[2025-11-08 17:50:20,794] INFO [RaftManager id=1] Attempting durable transition to UnattachedState(epoch=0, leaderId=OptionalInt.empty, votedKey=Optional.empty, voters=[1], electionTimeoutMs=1623, highWatermark=Optional.empty) from null (org.apache.kafka.raft.QuorumState)
[2025-11-08 17:50:20,805] INFO [RaftManager id=1] Completed transition to UnattachedState(epoch=0, leaderId=OptionalInt.empty, votedKey=Optional.empty, voters=[1], electionTimeoutMs=1623, highWatermark=Optional.empty) from null (org.apache.kafka.raft.QuorumState)
[2025-11-08 17:50:20,806] INFO [RaftManager id=1] Completed transition to ProspectiveState(epoch=0, leaderId=OptionalInt.empty, votedKey=Optional.empty, epochElection=EpochElection(voterStates={1=VoterState(replicaKey=ReplicaKey(id=1, directoryId=<undefined>), state=GRANTED)}), electionTimeoutMs=1929, highWatermark=Optional.empty) from UnattachedState(epoch=0, leaderId=OptionalInt.empty, votedKey=Optional.empty, voters=[1], electionTimeoutMs=1623, highWatermark=Optional.empty) (org.apache.kafka.raft.QuorumState)
[2025-11-08 17:50:20,807] INFO [RaftManager id=1] Attempting durable transition to CandidateState(localId=1, localDirectoryId=_QaHqb2s1M_RbZEtzGO0KQ, epoch=1, epochElection=EpochElection(voterStates={1=VoterState(replicaKey=ReplicaKey(id=1, directoryId=<undefined>), state=GRANTED)}), highWatermark=Optional.empty, electionTimeoutMs=1254) from ProspectiveState(epoch=0, leaderId=OptionalInt.empty, votedKey=Optional.empty, epochElection=EpochElection(voterStates={1=VoterState(replicaKey=ReplicaKey(id=1, directoryId=<undefined>), state=GRANTED)}), electionTimeoutMs=1929, highWatermark=Optional.empty) (org.apache.kafka.raft.QuorumState)
[2025-11-08 17:50:20,811] INFO [RaftManager id=1] Completed transition to CandidateState(localId=1, localDirectoryId=_QaHqb2s1M_RbZEtzGO0KQ, epoch=1, epochElection=EpochElection(voterStates={1=VoterState(replicaKey=ReplicaKey(id=1, directoryId=<undefined>), state=GRANTED)}), highWatermark=Optional.empty, electionTimeoutMs=1254) from ProspectiveState(epoch=0, leaderId=OptionalInt.empty, votedKey=Optional.empty, epochElection=EpochElection(voterStates={1=VoterState(replicaKey=ReplicaKey(id=1, directoryId=<undefined>), state=GRANTED)}), electionTimeoutMs=1929, highWatermark=Optional.empty) (org.apache.kafka.raft.QuorumState)
[2025-11-08 17:50:20,813] INFO [RaftManager id=1] Attempting durable transition to Leader(localVoterNode=VoterNode(voterKey=ReplicaKey(id=1, directoryId=_QaHqb2s1M_RbZEtzGO0KQ), listeners=Endpoints(endpoints={ListenerName(CONTROLLER)=localhost/<unresolved>:9093}), supportedKRaftVersion=SupportedVersionRange[min_version:0, max_version:1]), epoch=1, epochStartOffset=0, highWatermark=Optional.empty, voterStates={1=ReplicaState(replicaKey=ReplicaKey(id=1, directoryId=<undefined>), endOffset=Optional.empty, lastFetchTimestamp=-1, lastCaughtUpTimestamp=-1, hasAcknowledgedLeader=true)}) from CandidateState(localId=1, localDirectoryId=_QaHqb2s1M_RbZEtzGO0KQ, epoch=1, epochElection=EpochElection(voterStates={1=VoterState(replicaKey=ReplicaKey(id=1, directoryId=<undefined>), state=GRANTED)}), highWatermark=Optional.empty, electionTimeoutMs=1254) (org.apache.kafka.raft.QuorumState)
[2025-11-08 17:50:20,816] INFO [RaftManager id=1] Completed transition to Leader(localVoterNode=VoterNode(voterKey=ReplicaKey(id=1, directoryId=_QaHqb2s1M_RbZEtzGO0KQ), listeners=Endpoints(endpoints={ListenerName(CONTROLLER)=localhost/<unresolved>:9093}), supportedKRaftVersion=SupportedVersionRange[min_version:0, max_version:1]), epoch=1, epochStartOffset=0, highWatermark=Optional.empty, voterStates={1=ReplicaState(replicaKey=ReplicaKey(id=1, directoryId=<undefined>), endOffset=Optional.empty, lastFetchTimestamp=-1, lastCaughtUpTimestamp=-1, hasAcknowledgedLeader=true)}) from CandidateState(localId=1, localDirectoryId=_QaHqb2s1M_RbZEtzGO0KQ, epoch=1, epochElection=EpochElection(voterStates={1=VoterState(replicaKey=ReplicaKey(id=1, directoryId=<undefined>), state=GRANTED)}), highWatermark=Optional.empty, electionTimeoutMs=1254) (org.apache.kafka.raft.QuorumState)
[2025-11-08 17:50:20,827] INFO [kafka-1-raft-outbound-request-thread]: Starting (org.apache.kafka.raft.KafkaNetworkChannel$SendThread)
[2025-11-08 17:50:20,827] INFO [kafka-1-raft-io-thread]: Starting (org.apache.kafka.raft.KafkaRaftClientDriver)
[2025-11-08 17:50:20,831] INFO [MetadataLoader id=1] initializeNewPublishers: The loader is still catching up because we have loaded up to offset -1, but the high water mark is 1 (org.apache.kafka.image.loader.MetadataLoader)
[2025-11-08 17:50:20,834] INFO [ControllerServer id=1] Waiting for controller quorum voters future (kafka.server.ControllerServer)
[2025-11-08 17:50:20,834] INFO [ControllerServer id=1] Finished waiting for controller quorum voters future (kafka.server.ControllerServer)
[2025-11-08 17:50:20,835] INFO [RaftManager id=1] High watermark set to LogOffsetMetadata(offset=1, metadata=Optional[(segmentBaseOffset=0,relativePositionInSegment=91)]) for the first time for epoch 1 based on indexOfHw 0 and voters [ReplicaState(replicaKey=ReplicaKey(id=1, directoryId=<undefined>), endOffset=Optional[LogOffsetMetadata(offset=1, metadata=Optional[(segmentBaseOffset=0,relativePositionInSegment=91)])], lastFetchTimestamp=-1, lastCaughtUpTimestamp=-1, hasAcknowledgedLeader=true)] (org.apache.kafka.raft.LeaderState)
[2025-11-08 17:50:20,845] INFO [RaftManager id=1] Registered the listener org.apache.kafka.image.loader.MetadataLoader@1524067152 (org.apache.kafka.raft.KafkaRaftClient)
[2025-11-08 17:50:20,845] INFO [RaftManager id=1] Setting the next offset of org.apache.kafka.image.loader.MetadataLoader@1524067152 to 0 since there are no snapshots (org.apache.kafka.raft.KafkaRaftClient)
[2025-11-08 17:50:20,846] INFO [MetadataLoader id=1] maybePublishMetadata(LOG_DELTA): The loader is still catching up because we have not loaded a controller record as of offset 0 and high water mark is 1 (org.apache.kafka.image.loader.MetadataLoader)
[2025-11-08 17:50:20,848] INFO [RaftManager id=1] Registered the listener org.apache.kafka.controller.QuorumController$QuorumMetaLogListener@817418900 (org.apache.kafka.raft.KafkaRaftClient)
[2025-11-08 17:50:20,848] INFO [RaftManager id=1] Setting the next offset of org.apache.kafka.controller.QuorumController$QuorumMetaLogListener@817418900 to 0 since there are no snapshots (org.apache.kafka.raft.Ka
```

```
        socket.receive.buffer.bytes = 102400
        socket.request.max.bytes = 104857600
        socket.send.buffer.bytes = 102400
        ssl.allow.dn.changes = false
        ssl.allow.san.changes = false
        ssl.cipher.suites = []
        ssl.client.auth = none
        ssl.enabled.protocols = [TLSv1.2, TLSv1.3]
        ssl.endpoint.identification.algorithm = https
        ssl.engine.factory.class = null
        ssl.key.password = null
        ssl.keymanager.algorithm = SunX509
        ssl.keystore.certificate.chain = null
        ssl.keystore.key = null
        ssl.keystore.location = null
        ssl.keystore.password = null
        ssl.keystore.type = JKS
        ssl.principal.mapping.rules = DEFAULT
        ssl.protocol = TLSv1.3
        ssl.provider = null
        ssl.secure.random.implementation = null
        ssl.trustmanager.algorithm = PKIX
        ssl.truststore.certificates = null
        ssl.truststore.location = null
        ssl.truststore.password = null
        ssl.truststore.type = JKS
        telemetry.max.bytes = 1048576
        transaction.abort.timed.out.transaction.cleanup.interval.ms = 10000
        transaction.max.timeout.ms = 900000
        transaction.partition.verification.enable = true
        transaction.remove.expired.transaction.cleanup.interval.ms = 3600000
        transaction.state.log.load.buffer.size = 5242880
        transaction.state.log.min.isr = 1
        transaction.state.log.num.partitions = 50
        transaction.state.log.replication.factor = 1
        transaction.state.log.segment.bytes = 104857600
        transaction.two.phase.commit.enable = false
        transactional.id.expiration.ms = 604800000
        unclean.leader.election.enable = false
        unclean.leader.election.interval.ms = 300000
        unstable.api.versions.enable = false
        unstable.feature.versions.enable = false
 (org.apache.kafka.common.config.AbstractConfig)
[2025-11-08 17:50:21,009] INFO [BrokerServer id=1] Waiting for the broker to be unfenced (kafka.server.BrokerServer)
[2025-11-08 17:50:21,041] INFO [BrokerLifecycleManager id=1] The broker has been unfenced. Transitioning from RECOVERY to RUNNING. (kafka.server.BrokerLifecycleManager)
[2025-11-08 17:50:21,044] INFO [BrokerServer id=1] Finished waiting for the broker to be unfenced (kafka.server.BrokerServer)
[2025-11-08 17:50:21,044] INFO authorizerStart completed for endpoint PLAINTEXT. Endpoint is now READY. (org.apache.kafka.server.network.EndpointReadyFutures)
[2025-11-08 17:50:21,044] INFO [SocketServer listenerType=BROKER, nodeId=1] Enabling request processing. (kafka.network.SocketServer)
[2025-11-08 17:50:21,044] INFO Awaiting socket connections on 0.0.0.0:9092. (kafka.network.DataPlaneAcceptor)
[2025-11-08 17:50:21,044] INFO [BrokerServer id=1] Waiting for all of the authorizer futures to be completed (kafka.server.BrokerServer)
[2025-11-08 17:50:21,045] INFO [BrokerServer id=1] Finished waiting for all of the authorizer futures to be completed (kafka.server.BrokerServer)
[2025-11-08 17:50:21,045] INFO [BrokerServer id=1] Waiting for all of the SocketServer Acceptors to be started (kafka.server.BrokerServer)
[2025-11-08 17:50:21,045] INFO [BrokerServer id=1] Finished waiting for all of the SocketServer Acceptors to be started (kafka.server.BrokerServer)
[2025-11-08 17:50:21,045] INFO [BrokerServer id=1] Transition from STARTING to STARTED (kafka.server.BrokerServer)
[2025-11-08 17:50:21,045] INFO Kafka version: 4.1.0 (org.apache.kafka.common.utils.AppInfoParser)
[2025-11-08 17:50:21,045] INFO Kafka commitId: 13f70256db3c994c (org.apache.kafka.common.utils.AppInfoParser)
[2025-11-08 17:50:21,045] INFO Kafka startTimeMs: 1762604421045 (org.apache.kafka.common.utils.AppInfoParser)
[2025-11-08 17:50:21,045] INFO [KafkaRaftServer nodeId=1] Kafka Server started (kafka.server.KafkaRaftServer)
[2025-11-08 17:58:06,225] INFO [ReplicaFetcherManager on broker 1] Removed fetcher for partitions Set(wiki-vote-0) (kafka.server.ReplicaFetcherManager)
[2025-11-08 17:58:06,232] INFO [LogLoader partition=wiki-vote-0, dir=/opt/homebrew/var/lib/kraft-combined-logs] Loading producer state till offset 0 (org.apache.kafka.storage.internals.log.UnifiedLog)
[2025-11-08 17:58:06,232] INFO Created log for partition wiki-vote-0 in /opt/homebrew/var/lib/kraft-combined-logs/wiki-vote-0 with properties {} (kafka.log.LogManager)
[2025-11-08 17:58:06,236] INFO [Partition wiki-vote-0 broker=1] No checkpointed highwatermark is found for partition wiki-vote-0 (kafka.cluster.Partition)
[2025-11-08 17:58:06,237] INFO [Partition wiki-vote-0 broker=1] Log loaded for partition wiki-vote-0 with initial high watermark 0 (kafka.cluster.Partition)
[2025-11-08 18:00:21,049] INFO [NodeToControllerChannelManager id=1 name=registration] Node 1 disconnected. (org.apache.kafka.clients.NetworkClient)
[2025-11-08 18:08:06,292] INFO [NodeToControllerChannelManager id=1 name=forwarding] Node 1 disconnected. (org.apache.kafka.clients.NetworkClient)
```

Start Kafka-server daemon, an inf. running process responsible for recieving data from producers, storing it on disk, serving it to the consumers on demand.

Now write producer consumer codes, producer with configurable delays to check effect on throughput.

```
(pypyenv) bhavychawla@Bhavys-MacBook-Pro ~ % python consumer_fault_tolerant.py
[INFO] Connecting to Kafka topic: wiki-vote...
[✅] Consumer connected, waiting for data...
[🗂️] No checkpoint found, starting fresh graph.

[📈] Edges processed: 10000/103689 | Nodes: 1984
[⚡ Throughput: 2467.01 edges/sec (recent), 2467.01 avg | ⏳ ETA: 38.0 sec]

[📸 PARTIAL SNAPSHOT @ 10000 edges]
Nodes: 1984
Largest WCC: 1984 nodes, 10000 edges
Largest SCC: 95 nodes, 629 edges
Clustering coeff: 0.0523
Triangles: 215 | Closed triangle fraction: 0.02652
Diameter (sampled): 7 | Effective diameter (approx): 5.00

[📈] Edges processed: 20000/103689 | Nodes: 2331
[⚡ Throughput: 10967.88 edges/sec (recent), 4028.00 avg | ⏳ ETA: 20.8 sec]

[📸 PARTIAL SNAPSHOT @ 20000 edges]
Nodes: 2331
Largest WCC: 2331 nodes, 20000 edges
Largest SCC: 213 nodes, 2915 edges
Clustering coeff: 0.1798
Triangles: 1798 | Closed triangle fraction: 0.09391
Diameter (sampled): 6 | Effective diameter (approx): 4.00

[📈] Edges processed: 30000/103689 | Nodes: 2834
[⚡ Throughput: 8887.87 edges/sec (recent), 4925.81 avg | ⏳ ETA: 15.0 sec]

[📸 PARTIAL SNAPSHOT @ 30000 edges]
Nodes: 2834
Largest WCC: 2832 nodes, 29999 edges
Largest SCC: 371 nodes, 6498 edges
Clustering coeff: 0.0495
Triangles: 270 | Closed triangle fraction: 0.05559
Diameter (sampled): 7 | Effective diameter (approx): 4.00

[📈] Edges processed: 40000/103689 | Nodes: 3257
[⚡ Throughput: 1659.42 edges/sec (recent), 3301.27 avg | ⏳ ETA: 19.3 sec]

[📸 PARTIAL SNAPSHOT @ 40000 edges]
Nodes: 3257
Largest WCC: 3255 nodes, 39999 edges
Largest SCC: 447 nodes, 8848 edges
Clustering coeff: 0.0931
Triangles: 760 | Closed triangle fraction: 0.08650
Diameter (sampled): 6 | Effective diameter (approx): 4.00
[💾] Graph checkpoint saved (47226 edges).

[📈] Edges processed: 50000/103689 | Nodes: 3613
[⚡ Throughput: 402.99 edges/sec (recent), 1353.87 avg | ⏳ ETA: 39.7 sec]

[📸 PARTIAL SNAPSHOT @ 50000 edges]
Nodes: 3613
Largest WCC: 3611 nodes, 49999 edges
Largest SCC: 565 nodes, 12218 edges
Clustering coeff: 0.1095
Triangles: 1097 | Closed triangle fraction: 0.07628
Diameter (sampled): 6 | Effective diameter (approx): 4.00
[💾] Graph checkpoint saved (50858 edges).
[💾] Graph checkpoint saved (50858 edges).
^C
[⚠️ Interrupted manually. Saving checkpoint...
[🏁 Completed in 109.89 seconds total]
```

totally fault tolerant, it creates a checkpoint file one a json and other pkl, pkl to dump offset and json to dump the partial output.

```
========== FINAL METRICS ==========
nodes                    : 3631  (GT=7115)
edges                    : 50858  (GT=103689)
largest_wcc_nodes        : 3629  (GT=7066)
largest_wcc_edges        : 50857  (GT=103663)
largest_scc_nodes        : 580  (GT=1300)
largest_scc_edges        : 12835  (GT=39456)
clustering_coeff         : 0.15780532976294223  (GT=0.1409)
triangles                : 211451  (GT=608389)
closed_triangle_fraction : 0.09890218856014156  (GT=0.04564)
diameter                 : 6  (GT=7)
eff_diameter             : 5  (GT=3.8)
====================================
(pypyenv) bhavychawla@Bhavys-MacBook-Pro ~ % python consumer_fault_tolerant.py
[INFO] Connecting to Kafka topic: wiki-vote...
[✅] Consumer connected, waiting for data...
[🔄] Loaded checkpoint graph with 50858 edges.
[💾] Graph checkpoint saved (56659 edges).

[📈] Edges processed: 60000/103689 | Nodes: 3937
[⚡ Throughput: 258.44 edges/sec (recent), 1550.63 avg | ⏳ ETA: 28.2 sec]

[📸 PARTIAL SNAPSHOT @ 60000 edges]
Nodes: 3937
Largest WCC: 3935 nodes, 59999 edges
Largest SCC: 688 nodes, 17369 edges
Clustering coeff: 0.1080
Triangles: 1064 | Closed triangle fraction: 0.10522
Diameter (sampled): 7 | Effective diameter (approx): 4.00
[💾] Graph checkpoint saved (68461 edges).

[📈] Edges processed: 70000/103689 | Nodes: 4374
[⚡ Throughput: 393.75 edges/sec (recent), 1092.20 avg | ⏳ ETA: 30.8 sec]

[📸 PARTIAL SNAPSHOT @ 70000 edges]
Nodes: 4374
Largest WCC: 4368 nodes, 69997 edges
Largest SCC: 787 nodes, 21605 edges
Clustering coeff: 0.0884
Triangles: 955 | Closed triangle fraction: 0.11069
Diameter (sampled): 7 | Effective diameter (approx): 4.00

[📈] Edges processed: 80000/103689 | Nodes: 5074
[⚡ Throughput: 385.43 edges/sec (recent), 888.53 avg | ⏳ ETA: 26.7 sec]

[📸 PARTIAL SNAPSHOT @ 80000 edges]
Nodes: 5074
Largest WCC: 5064 nodes, 79995 edges
Largest SCC: 933 nodes, 26819 edges
Clustering coeff: 0.0503
Triangles: 381 | Closed triangle fraction: 0.11082
Diameter (sampled): 6 | Effective diameter (approx): 4.00
[💾] Graph checkpoint saved (80000 edges).

[📈] Edges processed: 90000/103689 | Nodes: 5645
[⚡ Throughput: 388.78 edges/sec (recent), 777.48 avg | ⏳ ETA: 17.6 sec]

[📸 PARTIAL SNAPSHOT @ 90000 edges]
Nodes: 5645
Largest WCC: 5631 nodes, 89993 edges
Largest SCC: 1067 nodes, 31783 edges
Clustering coeff: 0.0688
Triangles: 507 | Closed triangle fraction: 0.10897
Diameter (sampled): 6 | Effective diameter (approx): 4.00
[💾] Graph checkpoint saved (91976 edges).
```

→ outputs are partially reported after every 10000 edges.

```
[📸 PARTIAL SNAPSHOT @ 90000 edges]
Nodes: 5645
Largest WCC: 5631 nodes, 89993 edges
Largest SCC: 1067 nodes, 31783 edges
Clustering coeff: 0.0688
Triangles: 507 | Closed triangle fraction: 0.10897
Diameter (sampled): 6 | Effective diameter (approx): 4.00
[💾] Graph checkpoint saved (91976 edges).

[📈] Edges processed: 100000/103689 | Nodes: 6418
[⚡ Throughput: 399.05 edges/sec (recent), 710.14 avg | ⏳ ETA: 5.2 sec]

[📸 PARTIAL SNAPSHOT @ 100000 edges]
Nodes: 6418
Largest WCC: 6387 nodes, 99984 edges
Largest SCC: 1223 nodes, 37554 edges
Clustering coeff: 0.0626
Triangles: 653 | Closed triangle fraction: 0.14557
Diameter (sampled): 6 | Effective diameter (approx): 4.00

[✅] Full stream received, computing final metrics...

[🏁 Completed in 154.84 seconds total]

========== FINAL METRICS ==========
nodes                    : 7115  (GT=7115)
edges                    : 103689  (GT=103689)
largest_wcc_nodes        : 7066  (GT=7066)
largest_wcc_edges        : 103663  (GT=103663)
largest_scc_nodes        : 1300  (GT=1300)
largest_scc_edges        : 39456  (GT=39456)
clustering_coeff         : 0.14089784589308763  (GT=0.1409)
triangles                : 608389  (GT=608389)
closed_triangle_fraction : 0.12547914899233995  (GT=0.04564)
diameter                 : 7  (GT=7)
eff_diameter             : 4  (GT=3.8)
====================================
(pypyenv) bhavychawla@Bhavys-MacBook-Pro ~ %
```

} final metrics

total time ≈ 263 s. with delay 0.002s. ↓ producer side.

```
========== FINAL METRICS ==========
nodes                          : 7115  (GT=7115)
edges                          : 103689  (GT=103689)
largest_wcc_nodes              : 7066  (GT=7066)
largest_wcc_edges              : 103663  (GT=103663)
largest_scc_nodes              : 1300  (GT=1300)
largest_scc_edges              : 39456  (GT=39456)
clustering_coeff               : 0.14089784589308738  (GT=0.1409)
triangles                      : 608389  (GT=608389)
closed_triangle_fraction       : 0.12547914899233995  (GT=0.04564)
diameter                       : 7  (GT=7)
eff_diameter                   : 5  (GT=3.8)
=====================================
(pypyenv) bhavychawla@Bhavys-MacBook-Pro ~ % python consumer_fault_tolerant.py
[INFO] Connecting to Kafka topic: wiki-vote...
[↻] Producer restarted — clearing old consumer checkpoint for fresh start.
[✅] Consumer connected, waiting for data...
[🆕] No checkpoint found, starting fresh graph.
```

↓

once completed the producer consumer also restart with line number as 0.

line number → (src dst) pair.


```
========== FINAL METRICS ==========
nodes                          : 7115  (GT=7115)
edges                          : 103689  (GT=103689)
largest_wcc_nodes              : 7066  (GT=7066)
largest_wcc_edges              : 103663  (GT=103663)
largest_scc_nodes              : 1300  (GT=1300)
largest_scc_edges              : 39456  (GT=39456)
clustering_coeff               : 0.14089784589308738  (GT=0.1409)
triangles                      : 608389  (GT=608389)
closed_triangle_fraction       : 0.12547914899233995  (GT=0.04564)
diameter                       : 7  (GT=7)
eff_diameter                   : 5  (GT=3.8)
=====================================
(pypyenv) bhavychawla@Bhavys-MacBook-Pro ~ % python consumer_fault_tolerant.py
[INFO] Connecting to Kafka topic: wiki-vote...
[↻] Producer restarted — clearing old consumer checkpoint for fresh start.
[✅] Consumer connected, waiting for data...
[🆕] No checkpoint found, starting fresh graph.
```


↓

clears old checkpoints after computation is complete, when we start again automatically starts from line 0.

Comparision of throughput and completion time with diff producer delays:

delay = 0.002 s

```python
# producer_fault_tolerant.py
from kafka import KafkaProducer
import time, os, sys, json

TOPIC = "wiki-vote"
BOOTSTRAP_SERVERS = "localhost:9092"

# 🗂️ Update this path if your dataset is elsewhere
DATA_PATH = "/Users/bhavychawla/Downloads/Wiki-Vote.txt"
CHECKPOINT_FILE = os.path.abspath("producer_checkpoint.json")

# ⚙️ Tune this for speed vs stability
DELAY = 0.002  # seconds between sends

# ------------------------------------------------------------------
def create_producer():
    """Create a resilient Kafka producer with retries and acks."""
    try:
        producer = KafkaProducer(
            bootstrap_servers=BOOTSTRAP_SERVERS,
            acks="all",
            retries=15,
            linger_ms=5,
            max_in_flight_requests_per_connection=1,
        )
        print(f"[✅] Kafka Producer connected → Topic: {TOPIC}")
```

```
        return producer
    except Exception as e:
```

```
[🏁] Stream complete — checkpoint reset.
(pypyenv) bhavychawla@Bhavys-MacBook-Pro ~ % python producer_fault_tolerant.py
[✅] Kafka Producer connected → Topic: wiki-vote
[🔄] Resuming from line 0...
[STREAM] Sent 5000 edges... (line 5003)
[STREAM] Sent 10000 edges... (line 10003)
[STREAM] Sent 15000 edges... (line 15003)
^C
[⚠️] Producer interrupted manually, saving progress...
(pypyenv) bhavychawla@Bhavys-MacBook-Pro ~ % python producer_fault_tolerant.py
[✅] Kafka Producer connected → Topic: wiki-vote
[🔄] Resuming from line 17464...
[STREAM] Sent 5000 edges... (line 22463)
[STREAM] Sent 10000 edges... (line 27463)
[STREAM] Sent 15000 edges... (line 32463)
[STREAM] Sent 20000 edges... (line 37463)
[STREAM] Sent 25000 edges... (line 42463)
[STREAM] Sent 30000 edges... (line 47463)
```

```
eff_diameter               : 5  (CI=5.0)
====================================
(pypyenv) bhavychawla@Bhavys-MacBook-Pro ~ % python consumer_fault_tolerant.py
[INFO] Connecting to Kafka topic: wiki-vote...
[✅] Consumer connected, waiting for data...
[🔄] Loaded checkpoint graph with 16217 edges.

[📈] Edges processed: 20000/103689 | Nodes: 2438
[⚡ Throughput: 1815.23 edges/sec (recent), 3630.45 avg | ⏳ ETA: 23.1 sec]

[⏱ PARTIAL SNAPSHOT @ 20000 edges]
Nodes: 2438
Largest WCC: 2438 nodes, 20000 edges
Largest SCC: 178 nodes, 2373 edges
Clustering coeff: 0.0902
Triangles: 834 | Closed triangle fraction: 0.09835
Diameter (sampled): 5 | Effective diameter (approx): 4.00
[💾] Graph checkpoint saved (29234 edges).

[📈] Edges processed: 30000/103689 | Nodes: 2776
[⚡ Throughput: 377.95 edges/sec (recent), 938.44 avg | ⏳ ETA: 78.5 sec]

[⏱ PARTIAL SNAPSHOT @ 30000 edges]
Nodes: 2776
Largest WCC: 2776 nodes, 30000 edges
Largest SCC: 304 nodes, 4884 edges
Clustering coeff: 0.1241
Triangles: 778 | Closed triangle fraction: 0.04550
Diameter (sampled): 5 | Effective diameter (approx): 4.00

[📈] Edges processed: 40000/103689 | Nodes: 3123
[⚡ Throughput: 380.84 edges/sec (recent), 686.98 avg | ⏳ ETA: 92.7 sec]

[⏱ PARTIAL SNAPSHOT @ 40000 edges]
Nodes: 3123
Largest WCC: 3123 nodes, 40000 edges
Largest SCC: 436 nodes, 8441 edges
Clustering coeff: 0.1198
```

```
(pypyenv) bhavychawla@Bhavys-MacBook-Pro ~ % python producer_fault_tolerant.py
[✅] Kafka Producer connected → Topic: wiki-vote
[🔄] Resuming from line 0...
[STREAM] Sent 5000 edges... (line 5003)
[STREAM] Sent 10000 edges... (line 10003)
[STREAM] Sent 15000 edges... (line 15003)
[STREAM] Sent 20000 edges... (line 20003)
[STREAM] Sent 25000 edges... (line 25003)
[STREAM] Sent 30000 edges... (line 30003)
[STREAM] Sent 35000 edges... (line 35003)
[STREAM] Sent 40000 edges... (line 40003)
[STREAM] Sent 45000 edges... (line 45003)
[STREAM] Sent 50000 edges... (line 50003)
[STREAM] Sent 55000 edges... (line 55003)
[STREAM] Sent 60000 edges... (line 60003)
[STREAM] Sent 65000 edges... (line 65003)
[STREAM] Sent 70000 edges... (line 70003)
[STREAM] Sent 75000 edges... (line 75003)
[STREAM] Sent 80000 edges... (line 80003)
[STREAM] Sent 85000 edges... (line 85003)
[STREAM] Sent 90000 edges... (line 90003)
[STREAM] Sent 95000 edges... (line 95003)
[STREAM] Sent 100000 edges... (line 100003)

[✅] Streaming complete — Total edges sent: 103689
[🏁] Stream complete — checkpoint reset.
(pypyenv) bhavychawla@Bhavys-MacBook-Pro ~ % python producer_fault_tolerant.py
[✅] Kafka Producer connected → Topic: wiki-vote
[🔄] Resuming from line 0...
[STREAM] Sent 5000 edges... (line 5003)
[STREAM] Sent 10000 edges... (line 10003)
[STREAM] Sent 15000 edges... (line 15003)
^C
[⚠️] Producer interrupted manually, saving progress...
(pypyenv) bhavychawla@Bhavys-MacBook-Pro ~ % python producer_fault_tolerant.py
[✅] Kafka Producer connected → Topic: wiki-vote
[🔄] Resuming from line 17464...
[STREAM] Sent 5000 edges... (line 22463)
[STREAM] Sent 10000 edges... (line 27463)
[STREAM] Sent 15000 edges... (line 32463)
[STREAM] Sent 20000 edges... (line 37463)
[STREAM] Sent 25000 edges... (line 42463)
[STREAM] Sent 30000 edges... (line 47463)
[STREAM] Sent 35000 edges... (line 52463)
[STREAM] Sent 40000 edges... (line 57463)
[STREAM] Sent 45000 edges... (line 62463)
[STREAM] Sent 50000 edges... (line 67463)
[STREAM] Sent 55000 edges... (line 72463)
[STREAM] Sent 60000 edges... (line 77463)
[STREAM] Sent 65000 edges... (line 82463)
[STREAM] Sent 70000 edges... (line 87463)
[STREAM] Sent 75000 edges... (line 92463)
[STREAM] Sent 80000 edges... (line 97463)
[STREAM] Sent 85000 edges... (line 102463)

[✅] Streaming complete — Total edges sent: 86229
[🏁] Stream complete — checkpoint reset.
```

```
[📈] Edges processed: 70000/103689 | Nodes: 4374
[⚡ Throughput: 385.29 edges/sec (recent), 514.41 avg | ⏳ ETA: 65.5 sec]

[⏱ PARTIAL SNAPSHOT @ 70000 edges]
Nodes: 4374
Largest WCC: 4368 nodes, 69997 edges
Largest SCC: 787 nodes, 21605 edges
Clustering coeff: 0.0630
Triangles: 421 | Closed triangle fraction: 0.08863
Diameter (sampled): 6 | Effective diameter (approx): 4.00
[💾] Graph checkpoint saved (75678 edges).

[📈] Edges processed: 80000/103689 | Nodes: 5074
[⚡ Throughput: 396.25 edges/sec (recent), 495.92 avg | ⏳ ETA: 47.8 sec]

[⏱ PARTIAL SNAPSHOT @ 80000 edges]
Nodes: 5074
Largest WCC: 5064 nodes, 79995 edges
Largest SCC: 933 nodes, 26819 edges
Clustering coeff: 0.0545
Triangles: 389 | Closed triangle fraction: 0.10799
Diameter (sampled): 6 | Effective diameter (approx): 4.00
[💾] Graph checkpoint saved (87597 edges).

[📈] Edges processed: 90000/103689 | Nodes: 5645
[⚡ Throughput: 393.48 edges/sec (recent), 481.98 avg | ⏳ ETA: 28.4 sec]

[⏱ PARTIAL SNAPSHOT @ 90000 edges]
Nodes: 5645
Largest WCC: 5631 nodes, 89993 edges
Largest SCC: 1067 nodes, 31783 edges
Clustering coeff: 0.0718
Triangles: 580 | Closed triangle fraction: 0.11074
Diameter (sampled): 6 | Effective diameter (approx): 4.00
[💾] Graph checkpoint saved (99447 edges).

[📈] Edges processed: 100000/103689 | Nodes: 6418
[⚡ Throughput: 393.34 edges/sec (recent), 471.36 avg | ⏳ ETA: 7.8 sec]

[⏱ PARTIAL SNAPSHOT @ 100000 edges]
Nodes: 6418
Largest WCC: 6387 nodes, 99984 edges
Largest SCC: 1223 nodes, 37554 edges
Clustering coeff: 0.0749
Triangles: 771 | Closed triangle fraction: 0.13002
Diameter (sampled): 6 | Effective diameter (approx): 4.00

[✅] Full stream received, computing final metrics...

[🏁 Completed in 226.89 seconds total]

========== FINAL METRICS ==========
nodes                     : 7115  (GT=7115)
edges                     : 103689  (GT=103689)
largest_wcc_nodes         : 7066  (GT=7066)
largest_wcc_edges         : 103663  (GT=103663)
largest_scc_nodes         : 1300  (GT=1300)
largest_scc_edges         : 39456  (GT=39456)
clustering_coeff          : 0.14089784589308738  (GT=0.1409)
triangles                 : 608389  (GT=608389)
closed_triangle_fraction  : 0.12547914899233995  (GT=0.04564)
diameter                  : 7  (GT=7)
eff_diameter              : 4  (GT=3.8)
====================================
```

Completion time — 226.89 seconds.

```python
# producer_fault_tolerant.py
from kafka import KafkaProducer
import time, os, sys, json

TOPIC = "wiki-vote"
BOOTSTRAP_SERVERS = "localhost:9092"

# 🗂️ Update this path if your dataset is elsewhere
DATA_PATH = "/Users/bhavychawla/Downloads/Wiki-Vote.txt"
CHECKPOINT_FILE = os.path.abspath("producer_checkpoint.json")

# ⚙️ Tune this for speed vs stability
DELAY = 0.002 # seconds between sends

# ------------------------------------------------------------------
def create_producer():
    """Create a resilient Kafka producer with retries and acks."""
    try:
        producer = KafkaProducer(
            bootstrap_servers=BOOTSTRAP_SERVERS,
            acks="all",
            retries=15,
            linger_ms=5,
            max_in_flight_requests_per_connection=1,
        )
        print(f"[✅] Kafka Producer connected → Topic: {TOPIC}")
        return producer
    except Exception as e:
```

```
==================================
(pypyenv) bhavychawla@Bhavys-MacBook-Pro ~ % python consumer_fault_tolerant.py
[INFO] Connecting to Kafka topic: wiki-vote...
[🔄] Producer restarted — clearing old consumer checkpoint for fresh start.
[✅] Consumer connected, waiting for data...
[ℹ️] No checkpoint found, starting fresh graph.
[💾] Graph checkpoint saved (2534 edges).
[💾] Graph checkpoint saved (4966 edges).
[💾] Graph checkpoint saved (7434 edges).
[💾] Graph checkpoint saved (9909 edges).

[📈] Edges processed: 10000/103689 | Nodes: 1825
[⚡ Throughput: 82.43 edges/sec (recent), 82.43 avg | ⏳ ETA: 1136.6 sec]

[⏱ PARTIAL SNAPSHOT @ 10000 edges]
Nodes: 1825
Largest WCC: 1825 nodes, 10000 edges
Largest SCC: 69 nodes, 625 edges
Clustering coeff: 0.0472
Triangles: 171 | Closed triangle fraction: 0.03147
Diameter (sampled): 6 | Effective diameter (approx): 4.00
[💾] Graph checkpoint saved (12382 edges).
[💾] Graph checkpoint saved (14858 edges).
[💾] Graph checkpoint saved (17337 edges).
[💾] Graph checkpoint saved (19805 edges).

[📈] Edges processed: 20000/103689 | Nodes: 2438
[⚡ Throughput: 82.28 edges/sec (recent), 82.35 avg | ⏳ ETA: 1016.2 sec]

[⏱ PARTIAL SNAPSHOT @ 20000 edges]
Nodes: 2438
Largest WCC: 2438 nodes, 20000 edges
Largest SCC: 178 nodes, 2373 edges
Clustering coeff: 0.0670
Triangles: 487 | Closed triangle fraction: 0.03767
Diameter (sampled): 5 | Effective diameter (approx): 4.00
[💾] Graph checkpoint saved (22275 edges).
[💾] Graph checkpoint saved (24762 edges).
[💾] Graph checkpoint saved (27237 edges).
[💾] Graph checkpoint saved (29716 edges).

[📈] Edges processed: 30000/103689 | Nodes: 2776
[⚡ Throughput: 82.40 edges/sec (recent), 82.37 avg | ⏳ ETA: 894.6 sec]

[⏱ PARTIAL SNAPSHOT @ 30000 edges]
Nodes: 2776
Largest WCC: 2776 nodes, 30000 edges
Largest SCC: 304 nodes, 4884 edges
Clustering coeff: 0.0977
Triangles: 679 | Closed triangle fraction: 0.06501
Diameter (sampled): 6 | Effective diameter (approx): 4.00
[💾] Graph checkpoint saved (32188 edges).
[💾] Graph checkpoint saved (34666 edges).
[💾] Graph checkpoint saved (37133 edges).
[💾] Graph checkpoint saved (39588 edges).

[📈] Edges processed: 40000/103689 | Nodes: 3123
[⚡ Throughput: 82.00 edges/sec (recent), 82.28 avg | ⏳ ETA: 774.1 sec]

[⏱ PARTIAL SNAPSHOT @ 40000 edges]
Nodes: 3123
Largest WCC: 3123 nodes, 40000 edges
Largest SCC: 436 nodes, 8441 edges
Clustering coeff: 0.0978
Triangles: 631 | Closed triangle fraction: 0.05861
Diameter (sampled): 5 | Effective diameter (approx): 4.00
```

```
[⏱ PARTIAL SNAPSHOT @ 40000 edges]
Nodes: 3123
Largest WCC: 3123 nodes, 40000 edges
Largest SCC: 436 nodes, 8441 edges
Clustering coeff: 0.0978
Triangles: 631 | Closed triangle fraction: 0.05861
Diameter (sampled): 5 | Effective diameter (approx): 4.00
[💾] Graph checkpoint saved (42035 edges).
[💾] Graph checkpoint saved (44451 edges).
[💾] Graph checkpoint saved (46919 edges).
[💾] Graph checkpoint saved (49395 edges).

[📈] Edges processed: 50000/103689 | Nodes: 3620
[⚡ Throughput: 81.58 edges/sec (recent), 82.14 avg | ⏳ ETA: 653.6 sec]

[⏱ PARTIAL SNAPSHOT @ 50000 edges]
Nodes: 3620
Largest WCC: 3618 nodes, 49999 edges
Largest SCC: 576 nodes, 12520 edges
Clustering coeff: 0.1083
Triangles: 630 | Closed triangle fraction: 0.06760
Diameter (sampled): 6 | Effective diameter (approx): 4.00
[💾] Graph checkpoint saved (51869 edges).
[💾] Graph checkpoint saved (54349 edges).
[💾] Graph checkpoint saved (56822 edges).
[💾] Graph checkpoint saved (59299 edges).

[📈] Edges processed: 60000/103689 | Nodes: 3937
[⚡ Throughput: 82.26 edges/sec (recent), 82.16 avg | ⏳ ETA: 531.8 sec]

[⏱ PARTIAL SNAPSHOT @ 60000 edges]
Nodes: 3937
Largest WCC: 3935 nodes, 59999 edges
Largest SCC: 688 nodes, 17369 edges
Clustering coeff: 0.0732
Triangles: 772 | Closed triangle fraction: 0.12041
Diameter (sampled): 7 | Effective diameter (approx): 4.00
[💾] Graph checkpoint saved (61742 edges).
[💾] Graph checkpoint saved (64209 edges).
[💾] Graph checkpoint saved (66680 edges).
[💾] Graph checkpoint saved (69162 edges).

[📈] Edges processed: 70000/103689 | Nodes: 4374
[⚡ Throughput: 82.05 edges/sec (recent), 82.14 avg | ⏳ ETA: 410.1 sec]

[⏱ PARTIAL SNAPSHOT @ 70000 edges]
Nodes: 4374
Largest WCC: 4368 nodes, 69997 edges
Largest SCC: 787 nodes, 21605 edges
Clustering coeff: 0.0748
Triangles: 776 | Closed triangle fraction: 0.11583
Diameter (sampled): 7 | Effective diameter (approx): 4.00
[💾] Graph checkpoint saved (71637 edges).
[💾] Graph checkpoint saved (74125 edges).
[💾] Graph checkpoint saved (76596 edges).
[💾] Graph checkpoint saved (79082 edges).

[📈] Edges processed: 80000/103689 | Nodes: 5074
[⚡ Throughput: 82.31 edges/sec (recent), 82.16 avg | ⏳ ETA: 288.3 sec]

[⏱ PARTIAL SNAPSHOT @ 80000 edges]
Nodes: 5074
Largest WCC: 5064 nodes, 79995 edges
Largest SCC: 933 nodes, 26819 edges
Clustering coeff: 0.0523
```

```python
# producer_fault_tolerant.py
from kafka import KafkaProducer
import time, os, sys, json

TOPIC = "wiki-vote"
BOOTSTRAP_SERVERS = "localhost:9092"

# 🗂️ Update this path if your dataset is elsewhere
DATA_PATH = "/Users/bhavychawla/Downloads/Wiki-Vote.txt"
CHECKPOINT_FILE = os.path.abspath("producer_checkpoint.json")

# ⚙️ Tune this for speed vs stability
DELAY = 0.01  # seconds between sends

# ------------------------------------------------------------------
def create_producer():
    """Create a resilient Kafka producer with retries and acks."""
    try:
        producer = KafkaProducer(
            bootstrap_servers=BOOTSTRAP_SERVERS,
            acks="all",
            retries=15,
            linger_ms=5,
            max_in_flight_requests_per_connection=1,
        )
        print(f"[✅] Kafka Producer connected → Topic: {TOPIC}")
```

→ 0.01s

```
        return producer
    except Exception as e:
```

```
[⏲ PARTIAL SNAPSHOT @ 80000 edges]
Nodes: 5074
Largest WCC: 5064 nodes, 79995 edges
Largest SCC: 933 nodes, 26819 edges
Clustering coeff: 0.0523
Triangles: 338 | Closed triangle fraction: 0.11622
Diameter (sampled): 7 | Effective diameter (approx): 4.00
[💾] Graph checkpoint saved (81560 edges).
[💾] Graph checkpoint saved (84047 edges).
[💾] Graph checkpoint saved (86524 edges).
[💾] Graph checkpoint saved (89006 edges).

[📈] Edges processed: 90000/103689 | Nodes: 5645
[⚡ Throughput: 82.29 edges/sec (recent), 82.18 avg | ⏳ ETA: 166.6 sec]

[⏲ PARTIAL SNAPSHOT @ 90000 edges]
Nodes: 5645
Largest WCC: 5631 nodes, 89993 edges
Largest SCC: 1067 nodes, 31783 edges
Clustering coeff: 0.0482
Triangles: 409 | Closed triangle fraction: 0.10458
Diameter (sampled): 6 | Effective diameter (approx): 4.00
[💾] Graph checkpoint saved (91478 edges).
[💾] Graph checkpoint saved (93966 edges).
[💾] Graph checkpoint saved (96446 edges).
[💾] Graph checkpoint saved (98942 edges).

[📈] Edges processed: 100000/103689 | Nodes: 6418
[⚡ Throughput: 82.42 edges/sec (recent), 82.20 avg | ⏳ ETA: 44.9 sec]

[⏲ PARTIAL SNAPSHOT @ 100000 edges]
Nodes: 6418
Largest WCC: 6387 nodes, 99984 edges
Largest SCC: 1223 nodes, 37554 edges
Clustering coeff: 0.0664
Triangles: 419 | Closed triangle fraction: 0.12751
Diameter (sampled): 7 | Effective diameter (approx): 4.00
[💾] Graph checkpoint saved (101415 edges).

[✅] Full stream received, computing final metrics...

[🏁 Completed in 1266.78 seconds total]

========== FINAL METRICS ==========
nodes                       : 7115  (GT=7115)
edges                       : 103689  (GT=103689)
largest_wcc_nodes           : 7066  (GT=7066)
largest_wcc_edges           : 103663  (GT=103663)
largest_scc_nodes           : 1300  (GT=1300)
largest_scc_edges           : 39456  (GT=39456)
clustering_coeff            : 0.14089784589308738  (GT=0.1409)
triangles                   : 608389  (GT=608389)
closed_triangle_fraction    : 0.12547914899233995  (GT=0.04564)
diameter                    : 7  (GT=7)
eff_diameter                : 4  (GT=3.8)
====================================
```

Run with delay = 0.01 s

Completion time = 1266.78 seconds.

```
(pypyenv) bhavychawla@Bhavys-MacBook-Pro ~ % python consumer_fault_tolerant.py
[INFO] Connecting to Kafka topic: wiki-vote...
[↻ ] Producer restarted — clearing old consumer checkpoint for fresh start.
[✅] Consumer connected, waiting for data...
[📦] No checkpoint found, starting fresh graph.

[📈] Edges processed: 10000/103689 | Nodes: 1825
[⚡ Throughput: 396.60 edges/sec (recent), 396.60 avg | ⏳ ETA: 236.2 sec]

[⏲ PARTIAL SNAPSHOT @ 10000 edges]
Nodes: 1825
Largest WCC: 1825 nodes, 10000 edges
Largest SCC: 69 nodes, 625 edges
Clustering coeff: 0.0893
Triangles: 844 | Closed triangle fraction: 0.06016
Diameter (sampled): 5 | Effective diameter (approx): 5.00
[💾] Graph checkpoint saved (11798 edges).

[📈] Edges processed: 20000/103689 | Nodes: 2438
[⚡ Throughput: 379.71 edges/sec (recent), 387.97 avg | ⏳ ETA: 215.7 sec]

[⏲ PARTIAL SNAPSHOT @ 20000 edges]
Nodes: 2438
Largest WCC: 2438 nodes, 20000 edges
Largest SCC: 178 nodes, 2373 edges
Clustering coeff: 0.1036
Triangles: 793 | Closed triangle fraction: 0.04093
Diameter (sampled): 5 | Effective diameter (approx): 4.00
[💾] Graph checkpoint saved (23249 edges).

[📈] Edges processed: 30000/103689 | Nodes: 2776
[⚡ Throughput: 378.80 edges/sec (recent), 384.87 avg | ⏳ ETA: 191.5 sec]

[⏲ PARTIAL SNAPSHOT @ 30000 edges]
Nodes: 2776
Largest WCC: 2776 nodes, 30000 edges
Largest SCC: 304 nodes, 4884 edges
Clustering coeff: 0.1094
Triangles: 822 | Closed triangle fraction: 0.03522
Diameter (sampled): 5 | Effective diameter (approx): 5.00
[💾] Graph checkpoint saved (34633 edges).

[📈] Edges processed: 40000/103689 | Nodes: 3123
[⚡ Throughput: 379.14 edges/sec (recent), 383.42 avg | ⏳ ETA: 166.1 sec]

[⏲ PARTIAL SNAPSHOT @ 40000 edges]
Nodes: 3123
Largest WCC: 3123 nodes, 40000 edges
Largest SCC: 436 nodes, 8441 edges
Clustering coeff: 0.0950
Triangles: 1208 | Closed triangle fraction: 0.08576
Diameter (sampled): 5 | Effective diameter (approx): 4.00
[💾] Graph checkpoint saved (46041 edges).

[📈] Edges processed: 50000/103689 | Nodes: 3620
[⚡ Throughput: 382.03 edges/sec (recent), 383.14 avg | ⏳ ETA: 140.1 sec]

[⏲ PARTIAL SNAPSHOT @ 50000 edges]
Nodes: 3620
Largest WCC: 3618 nodes, 49999 edges
Largest SCC: 576 nodes, 12520 edges
Clustering coeff: 0.0473
Triangles: 309 | Closed triangle fraction: 0.07912
Diameter (sampled): 7 | Effective diameter (approx): 4.00
[💾] Graph checkpoint saved (57602 edges).
```

```
[⏲ PARTIAL SNAPSHOT @ 50000 edges]
Nodes: 3620
Largest WCC: 3618 nodes, 49999 edges
Largest SCC: 576 nodes, 12520 edges
Clustering coeff: 0.0473
Triangles: 309 | Closed triangle fraction: 0.07912
Diameter (sampled): 7 | Effective diameter (approx): 4.00
[💾] Graph checkpoint saved (57602 edges).

[📈] Edges processed: 60000/103689 | Nodes: 3937
[⚡ Throughput: 385.02 edges/sec (recent), 383.45 avg | ⌛ ETA: 113.9 sec]

[⏲ PARTIAL SNAPSHOT @ 60000 edges]
Nodes: 3937
Largest WCC: 3935 nodes, 59999 edges
Largest SCC: 688 nodes, 17369 edges
Clustering coeff: 0.0732
Triangles: 927 | Closed triangle fraction: 0.11700
Diameter (sampled): 7 | Effective diameter (approx): 4.00
[💾] Graph checkpoint saved (69354 edges).

[📈] Edges processed: 70000/103689 | Nodes: 4374
[⚡ Throughput: 391.38 edges/sec (recent), 384.56 avg | ⌛ ETA: 87.6 sec]

[⏲ PARTIAL SNAPSHOT @ 70000 edges]
Nodes: 4374
Largest WCC: 4368 nodes, 69997 edges
Largest SCC: 787 nodes, 21605 edges
Clustering coeff: 0.0753
Triangles: 675 | Closed triangle fraction: 0.09231
Diameter (sampled): 7 | Effective diameter (approx): 4.00

[📈] Edges processed: 80000/103689 | Nodes: 5074
[⚡ Throughput: 397.00 edges/sec (recent), 386.08 avg | ⌛ ETA: 61.4 sec]

[⏲ PARTIAL SNAPSHOT @ 80000 edges]
Nodes: 5074
Largest WCC: 5064 nodes, 79995 edges
Largest SCC: 933 nodes, 26819 edges
Clustering coeff: 0.0717
Triangles: 557 | Closed triangle fraction: 0.11265
Diameter (sampled): 6 | Effective diameter (approx): 4.00
[💾] Graph checkpoint saved (81315 edges).

[📈] Edges processed: 90000/103689 | Nodes: 5645
[⚡ Throughput: 400.24 edges/sec (recent), 387.60 avg | ⌛ ETA: 35.3 sec]

[⏲ PARTIAL SNAPSHOT @ 90000 edges]
Nodes: 5645
Largest WCC: 5631 nodes, 89993 edges
Largest SCC: 1067 nodes, 31783 edges
Clustering coeff: 0.0729
Triangles: 514 | Closed triangle fraction: 0.10849
Diameter (sampled): 6 | Effective diameter (approx): 4.00
[💾] Graph checkpoint saved (93389 edges).

[📈] Edges processed: 100000/103689 | Nodes: 6418
[⚡ Throughput: 405.87 edges/sec (recent), 389.35 avg | ⌛ ETA: 9.5 sec]

[⏲ PARTIAL SNAPSHOT @ 100000 edges]
Nodes: 6418
Largest WCC: 6387 nodes, 99984 edges
Largest SCC: 1223 nodes, 37554 edges
Clustering coeff: 0.0896
Triangles: 758 | Closed triangle fraction: 0.13568
```

```
[⏲ PARTIAL SNAPSHOT @ 100000 edges]
Nodes: 6418
Largest WCC: 6387 nodes, 99984 edges
Largest SCC: 1223 nodes, 37554 edges
Clustering coeff: 0.0896
Triangles: 758 | Closed triangle fraction: 0.13568
Diameter (sampled): 6 | Effective diameter (approx): 4.00

[✅] Full stream received, computing final metrics...

[※ Completed in 270.49 seconds total]

========== FINAL METRICS ==========
nodes                          : 7115  (GT=7115)
edges                          : 103689  (GT=103689)
largest_wcc_nodes              : 7066  (GT=7066)
largest_wcc_edges              : 103663  (GT=103663)
largest_scc_nodes              : 1300  (GT=1300)
largest_scc_edges              : 39456  (GT=39456)
clustering_coeff               : 0.14089784589388738  (GT=0.1409)
triangles                      : 608389  (GT=608389)
closed_triangle_fraction       : 0.12547914899233995  (GT=0.04564)
diameter                       : 7  (GT=7)
eff_diameter                   : 4  (GT=3.8)
====================================
```

Another run with delay = 0.002 s
Completion time = 270.49 s.

```python
# producer_fault_tolerant.py
from kafka import KafkaProducer
import time, os, sys, json

TOPIC = "wiki-vote"
BOOTSTRAP_SERVERS = "localhost:9092"

# 📂 Update this path if your dataset is elsewhere
DATA_PATH = "/Users/bhavychawla/Downloads/Wiki-Vote.txt"
CHECKPOINT_FILE = os.path.abspath("producer_checkpoint.json")

# ⚙️ Tune this for speed vs stability
DELAY = 0.0002  # seconds between sends

# ---------------------------------------------------------------------
def create_producer():
    """Create a resilient Kafka producer with retries and acks."""
    try:
        producer = KafkaProducer(
            bootstrap_servers=BOOTSTRAP_SERVERS,
            acks="all",
            retries=15,
            linger_ms=5,
            max_in_flight_requests_per_connection=1,
        )
        print(f"[✅] Kafka Producer connected → Topic: {TOPIC}")
        return producer
    except Exception as e:
        print(f"[❌] Kafka connection failed: {e}")
```

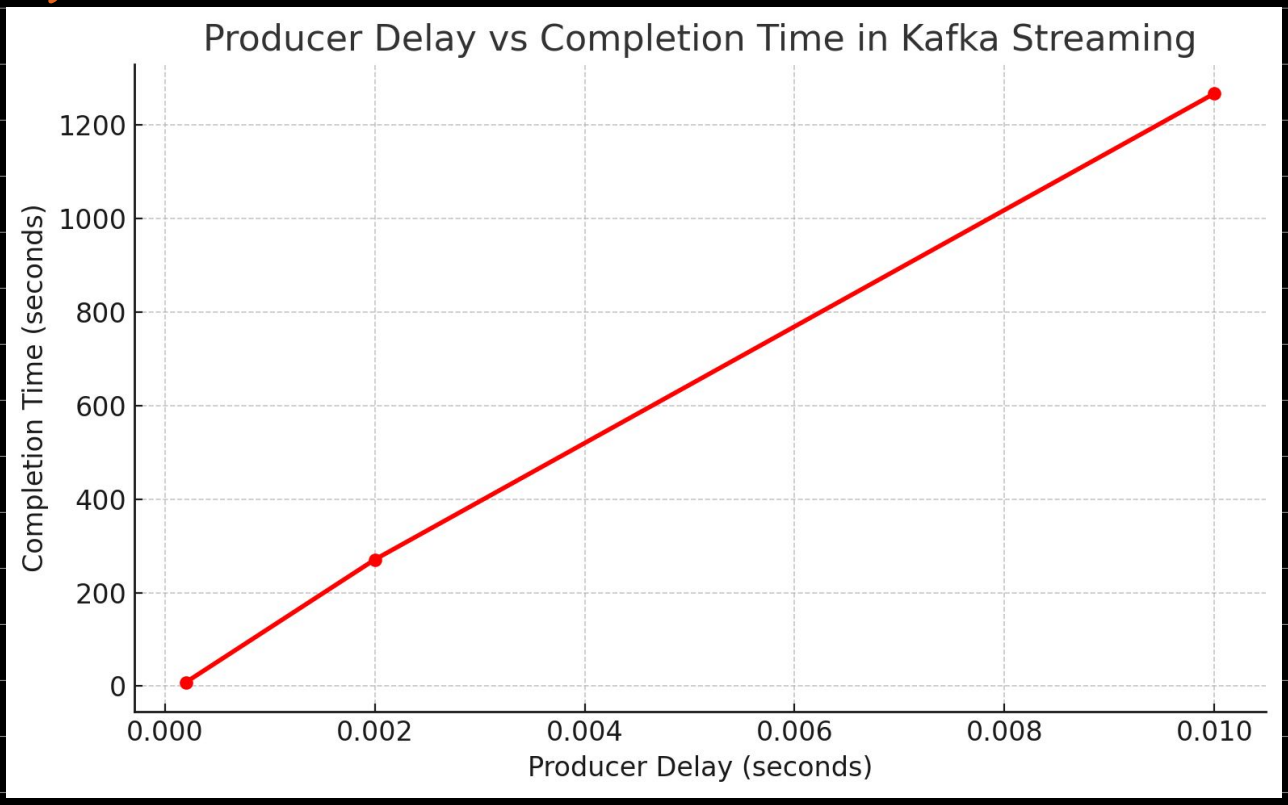
```
(pypyenv) bhavychawla@Bhavys-MacBook-Pro ~ % python producer_fault_tolerant.py
[✅] Kafka Producer connected → Topic: wiki-vote
[🔁] Resuming from line 0...
[STREAM] Sent 5000 edges... (line 5003)
[STREAM] Sent 10000 edges... (line 10003)
[STREAM] Sent 15000 edges... (line 15003)
[STREAM] Sent 20000 edges... (line 20003)
[STREAM] Sent 25000 edges... (line 25003)
[STREAM] Sent 30000 edges... (line 30003)
[STREAM] Sent 35000 edges... (line 35003)
[STREAM] Sent 40000 edges... (line 40003)
[STREAM] Sent 45000 edges... (line 45003)
[STREAM] Sent 50000 edges... (line 50003)
[STREAM] Sent 55000 edges... (line 55003)
[STREAM] Sent 60000 edges... (line 60003)
[STREAM] Sent 65000 edges... (line 65003)
[STREAM] Sent 70000 edges... (line 70003)
[STREAM] Sent 75000 edges... (line 75003)
[STREAM] Sent 80000 edges... (line 80003)
[STREAM] Sent 85000 edges... (line 85003)
[STREAM] Sent 90000 edges... (line 90003)
[STREAM] Sent 95000 edges... (line 95003)
[STREAM] Sent 100000 edges... (line 100003)

[✅] Streaming complete — Total edges sent: 103689
[🧹] Stream complete — checkpoint reset.
```

```
(pypyenv) bhavychawla@Bhavys-MacBook-Pro ~ % python consumer_fault_tolerant.py
[INFO] Connecting to Kafka topic: wiki-vote...
[✅] Consumer connected, waiting for data...
[🔁] Loaded checkpoint graph with 103689 edges.

[✅] Full stream received, computing final metrics...

[🏁 Completed in 8.18 seconds total]

========== FINAL METRICS ==========
nodes                   : 7115  (GT=7115)
edges                   : 103689  (GT=103689)
largest_wcc_nodes       : 7066  (GT=7066)
largest_wcc_edges       : 103663  (GT=103663)
largest_scc_nodes       : 1300  (GT=1300)
largest_scc_edges       : 39456  (GT=39456)
clustering_coeff        : 0.14089784589308738  (GT=0.1409)
triangles               : 608389  (GT=608389)
closed_triangle_fraction : 0.12547914899233995  (GT=0.04564)
diameter                : 7  (GT=7)
eff_diameter            : 4  (GT=3.8)
====================================
```

Completion time = 8.18s.

Conclusion: As delay increases throughput decreases, even in successive runs, and completion time increases.

final graph:


Producer Delay vs Completion Time in Kafka Streaming
Completion Time (seconds)
0
200
400
600
800
1000
1200
0.000
0.002
0.004
0.006
0.008
0.010
Producer Delay (seconds)