**kafka-topics --create --zookeeper ip-10-1-1-204.ap-south-1.compute.internal:2181 --replication-factor 1 --partitions 1 --topic raghavendr48\_test\_3005**

23/05/30 09:34:00 INFO zk.AdminZkClient: Topic creation Map(raghavendr48\_test\_3005-0 -> ArrayBuffer(183))

Created topic "raghavendr48\_test\_3005".

23/05/30 09:34:00 INFO zookeeper.ZooKeeperClient: [ZooKeeperClient] Closing.

23/05/30 09:34:00 INFO zookeeper.ZooKeeper: Session: 0xff87a02921053110 closed

23/05/30 09:34:00 INFO zookeeper.ClientCnxn: EventThread shut down

23/05/30 09:34:00 INFO zookeeper.ZooKeeperClient: [ZooKeeperClient] Closed.

**kafka-topics --describe --zookeeper ip-10-1-1-204.ap-south-1.compute.internal:2181 --topic raghavendr48\_test\_3005**

Topic:--raghavendr48\_test\_3005 PartitionCount:1 ReplicationFactor:1 Configs:

Topic: --raghavendr48\_test\_3005 Partition: 0 Leader: 183 Replicas: 183 Isr: 183

23/05/30 09:39:33 INFO zookeeper.ZooKeeperClient: [ZooKeeperClient] Closing.

23/05/30 09:39:33 INFO zookeeper.ZooKeeper: Session: 0xff87a02921053118 closed

23/05/30 09:39:33 INFO zookeeper.ClientCnxn: EventThread shut down

23/05/30 09:39:33 INFO zookeeper.ZooKeeperClient: [ZooKeeperClient] Closed.

**kafka-topics --list --zookeeper ip-10-1-1-204.ap-south-1.compute.internal:2181**

23/05/30 12:26:19 INFO zookeeper.ZooKeeperClient: [ZooKeeperClient] Connected.

--alabmanpreet

--edureka\_topic\_deepak

--kafkaTopic1

--raghavendr48\_test\_3005

00211telecom

11CaseStudy2Vin

123YAB

123YAB1

432544

8\_pranay

9CaseStudy2Vin

23/05/30 12:26:20 INFO zookeeper.ZooKeeperClient: [ZooKeeperClient] Closing.

23/05/30 12:26:20 INFO zookeeper.ZooKeeper: Session: 0xff87a02921053201 closed

23/05/30 12:26:20 INFO zookeeper.ClientCnxn: EventThread shut down

23/05/30 12:26:20 INFO zookeeper.ZooKeeperClient: [ZooKeeperClient] Closed.

[raghavendr48edu@ip-10-1-1-204 ~]$

**kafka-topics --list --zookeeper ip-10-1-1-204.ap-south-1.compute.internal:2181 | grep raghavendr48\_test\_\_3005**

If file not found

**0xff87a02921053211, negotiated timeout = 30000**

**23/05/30 12:36:22 INFO zookeeper.ZooKeeperClient: [ZooKeeperClient] Connected.**

**23/05/30 12:36:23 INFO zookeeper.ZooKeeperClient: [ZooKeeperClient] Closing.**

If file found

23/05/30 12:43:48 INFO zookeeper.ZooKeeperClient: [ZooKeeperClient] Connected.

--**raghavendr48\_test\_3005**

**raghavendr48\_test\_3005**

23/05/30 12:43:48 INFO zookeeper.ZooKeeperClient: [ZooKeeperClient] Closing.

23/05/30 12:43:48 INFO zookeeper.ZooKeeper: Session: 0xff87a0292105321f closed

23/05/30 12:43:48 INFO zookeeper.ClientCnxn: EventThread shut down

23/05/30 12:43:48 INFO zookeeper.ZooKeeperClient: [ZooKeeperClient] Closed.

[raghavendr48edu@ip-10-1-1-204 ~]$

**kafka-console-producer --broker-list ip-10-1-1-204.ap-south-1.compute.internal:9092 --topic raghavendr48\_test\_3005**

23/05/30 13:03:08 INFO utils.AppInfoParser: Kafka version : 2.1.0-cdh6.2.1

23/05/30 13:03:08 INFO utils.AppInfoParser: Kafka commitId : unknown

>hello

23/05/30 13:03:40 INFO clients.Metadata: Cluster ID: i6LyuYbbSMiPQc6ARDUuCg

>I

>am

>learning

>spark

>and

>kafka

>

**kafka-console-consumer --bootstrap-server ip-10-1-1-204.ap-south-1.compute.internal:9092 --topic raghavendr48\_test\_3005 --from-beginning**

23/05/30 13:11:12 INFO internals.Fetcher: [Consumer clientId=consumer-1, groupId=console-consumer-20698] Resetting offset for partition raghavendr48\_test\_300

5-0 to offset 0.

hello

I

am

learning

spark

and

kafka

>I

>love

>kafka

>flume

>and

>Spark

>

*From <*[*https://bigdata.cloudlabs.edureka.co/shell/*](https://bigdata.cloudlabs.edureka.co/shell/)*>*

vendr48\_test\_3005-0]

23/05/30 13:11:12 INFO internals.Fetcher: [Consumer clientId=consumer-1, groupId=console-consumer-20698] Resetting offset for partition raghavendr48\_test\_300

5-0 to offset 0.

hello

I

am

learning

spark

and

kafka

I

love

kafka

flume

and

Spark

*From <*[*https://bigdata.cloudlabs.edureka.co/shell/*](https://bigdata.cloudlabs.edureka.co/shell/)*>*

23/05/30 13:51:48 INFO utils.AppInfoParser: Kafka version : 2.1.0-cdh6.2.1

23/05/30 13:51:48 INFO utils.AppInfoParser: Kafka commitId : unknown

>I

23/05/30 13:51:59 INFO clients.Metadata: Cluster ID: i6LyuYbbSMiPQc6ARDUuCg

>love

>spark

>

*From <*[*https://bigdata.cloudlabs.edureka.co/shell/*](https://bigdata.cloudlabs.edureka.co/shell/)*>*

23/05/30 13:52:23 INFO internals.AbstractCoordinator: [Consumer clientId=consumer-1, groupId=console-consumer-74215] (Re-)joining group

23/05/30 13:52:26 INFO internals.AbstractCoordinator: [Consumer clientId=consumer-1, groupId=console-consumer-74215] Successfully joined group with generatio

n 1

23/05/30 13:52:26 INFO internals.ConsumerCoordinator: [Consumer clientId=consumer-1, groupId=console-consumer-74215] Setting newly assigned partitions [ragha

vendr48\_test\_3005-0]

23/05/30 13:52:26 INFO internals.Fetcher: [Consumer clientId=consumer-1, groupId=console-consumer-74215] Resetting offset for partition raghavendr48\_test\_300

5-0 to offset 0.

hello

I

am

learning

spark

and

kafka

I

love

kafka

flume

and

Spark

I

love

spark

*From <*[*https://bigdata.cloudlabs.edureka.co/shell/*](https://bigdata.cloudlabs.edureka.co/shell/)*>*

**kafka-topics --create --zookeeper ip-10-1-1-204.ap-south-1.compute.internal:2181 --replication-factor 1 --partitions 3 --topic raghavendr48\_test\_keyvalue\_3005**

23/05/30 13:22:03 INFO zookeeper.ZooKeeperClient: [ZooKeeperClient] Connected.

WARNING: Due to limitations in metric names, topics with a period ('.') or underscore ('\_') could collide. To avoid issues it is best to use either, but not

both.

23/05/30 13:22:03 INFO zk.AdminZkClient: Topic creation Map(raghavendr48\_test\_keyvalue\_3005-2 -> ArrayBuffer(184), raghavendr48\_test\_keyvalue\_3005-1 -> Array

Buffer(183), raghavendr48\_test\_keyvalue\_3005-0 -> ArrayBuffer(187))

Created topic "raghavendr48\_test\_keyvalue\_3005".

23/05/30 13:22:03 INFO zookeeper.ZooKeeperClient: [ZooKeeperClient] Closing.

23/05/30 13:22:03 INFO zookeeper.ZooKeeper: Session: 0xff87a02921053255 closed

23/05/30 13:22:03 INFO zookeeper.ClientCnxn: EventThread shut down

23/05/30 13:22:03 INFO zookeeper.ZooKeeperClient: [ZooKeeperClient] Closed.

*From <*[*https://bigdata.cloudlabs.edureka.co/shell/*](https://bigdata.cloudlabs.edureka.co/shell/)*>*

**kafka-topics --describe --zookeeper ip-10-1-1-204.ap-south-1.compute.internal:2181 --topic raghavendr48\_test\_keyvalue\_3005**

23/05/30 13:24:54 INFO zookeeper.ZooKeeperClient: [ZooKeeperClient] Connected.

Topic:raghavendr48\_test\_keyvalue\_3005 PartitionCount:3 ReplicationFactor:1 Configs:

Topic: raghavendr48\_test\_keyvalue\_3005 Partition: 0 Leader: 187 Replicas: 187 Isr: 187

Topic: raghavendr48\_test\_keyvalue\_3005 Partition: 1 Leader: 183 Replicas: 183 Isr: 183

Topic: raghavendr48\_test\_keyvalue\_3005 Partition: 2 Leader: 184 Replicas: 184 Isr: 184

23/05/30 13:24:54 INFO zookeeper.ZooKeeperClient: [ZooKeeperClient] Closing.

23/05/30 13:24:54 INFO zookeeper.ZooKeeper: Session: 0xff87a0292105325a closed

23/05/30 13:24:54 INFO zookeeper.ClientCnxn: EventThread shut down

23/05/30 13:24:54 INFO zookeeper.ZooKeeperClient: [ZooKeeperClient] Closed.

*From <*[*https://bigdata.cloudlabs.edureka.co/shell/*](https://bigdata.cloudlabs.edureka.co/shell/)*>*

**kafka-console-producer --broker-list ip-10-1-1-204.ap-south-1.compute.internal:9092 --topic raghavendr48\_test\_keyvalue\_3005 --property "parse.key=true" --property "key.separator=:"**

23/05/30 13:55:07 INFO utils.AppInfoParser: Kafka version : 2.1.0-cdh6.2.1

23/05/30 13:55:07 INFO utils.AppInfoParser: Kafka commitId : unknown

>1:spark

23/05/30 13:55:12 INFO clients.Metadata: Cluster ID: i6LyuYbbSMiPQc6ARDUuCg

>2:kafka

>3:flume

>4:hdfs

>

*From <*[*https://bigdata.cloudlabs.edureka.co/shell/*](https://bigdata.cloudlabs.edureka.co/shell/)*>*

**kafka-console-consumer --bootstrap-server ip-10-1-1-204.ap-south-1.compute.internal:9092 --topic raghavendr48\_test\_keyvalue\_3005 --from-beginning --partition 2**

23/05/30 13:59:14 INFO utils.AppInfoParser: Kafka version : 2.1.0-cdh6.2.1

23/05/30 13:59:14 INFO utils.AppInfoParser: Kafka commitId : unknown

23/05/30 13:59:14 INFO clients.Metadata: Cluster ID: i6LyuYbbSMiPQc6ARDUuCg

23/05/30 13:59:14 INFO internals.Fetcher: [Consumer clientId=consumer-1, groupId=console-consumer-32510] Resetting offset for partition raghavendr48\_test\_key

value\_3005-2 to offset 0.

kafka

flume

*From <*[*https://bigdata.cloudlabs.edureka.co/shell/*](https://bigdata.cloudlabs.edureka.co/shell/)*>*

**kafka-console-consumer --bootstrap-server ip-10-1-1-204.ap-south-1.compute.internal:9092 --topic raghavendr48\_test\_keyvalue\_3005 --from-beginning --partition 0**

23/05/30 14:02:09 INFO utils.AppInfoParser: Kafka version : 2.1.0-cdh6.2.1

23/05/30 14:02:09 INFO utils.AppInfoParser: Kafka commitId : unknown

23/05/30 14:02:09 INFO clients.Metadata: Cluster ID: i6LyuYbbSMiPQc6ARDUuCg

23/05/30 14:02:09 INFO internals.Fetcher: [Consumer clientId=consumer-1, groupId=console-consumer-99847] Resetting offset for partition raghavendr48\_test\_key

value\_3005-0 to offset 0.

spark

*From <*[*https://bigdata.cloudlabs.edureka.co/shell/*](https://bigdata.cloudlabs.edureka.co/shell/)*>*

**kafka-console-consumer --bootstrap-server ip-10-1-1-204.ap-south-1.compute.internal:9092 --topic raghavendr48\_test\_keyvalue\_3005 --from-beginning --partition 1**

23/05/30 14:02:48 INFO utils.AppInfoParser: Kafka version : 2.1.0-cdh6.2.1

23/05/30 14:02:48 INFO utils.AppInfoParser: Kafka commitId : unknown

23/05/30 14:02:48 INFO clients.Metadata: Cluster ID: i6LyuYbbSMiPQc6ARDUuCg

23/05/30 14:02:48 INFO internals.Fetcher: [Consumer clientId=consumer-1, groupId=console-consumer-88663] Resetting offset for partition raghavendr48\_test\_key

value\_3005-1 to offset 0.

hdfs

*From <*[*https://bigdata.cloudlabs.edureka.co/shell/*](https://bigdata.cloudlabs.edureka.co/shell/)*>*

23/05/30 13:55:07 INFO utils.AppInfoParser: Kafka version : 2.1.0-cdh6.2.1

23/05/30 13:55:07 INFO utils.AppInfoParser: Kafka commitId : unknown

>1:spark

23/05/30 13:55:12 INFO clients.Metadata: Cluster ID: i6LyuYbbSMiPQc6ARDUuCg

>2:kafka

>3:flume

>4:hdfs

>2:kafka\_new

>3:hdfs\_new

>4:flume\_new

>

3 new records added

*From <*[*https://bigdata.cloudlabs.edureka.co/shell/*](https://bigdata.cloudlabs.edureka.co/shell/)*>*

Without executing flume\_new came into parition 1

23/05/30 14:02:48 INFO utils.AppInfoParser: Kafka version : 2.1.0-cdh6.2.1

23/05/30 14:02:48 INFO utils.AppInfoParser: Kafka commitId : unknown

23/05/30 14:02:48 INFO clients.Metadata: Cluster ID: i6LyuYbbSMiPQc6ARDUuCg

23/05/30 14:02:48 INFO internals.Fetcher: [Consumer clientId=consumer-1, groupId=console-consumer-88663] Resetting offset for partition raghavendr48\_test\_key

value\_3005-1 to offset 0.

hdfs

flume\_new

*From <*[*https://bigdata.cloudlabs.edureka.co/shell/*](https://bigdata.cloudlabs.edureka.co/shell/)*>*

[raghavendr48edu@ip-10-1-1-204 ~]$ **kafka-console-producer --topic word\_count\_31052023 --broker-list ip-10-1-1-204.ap-south-1.compute.internal:9092, ip-10-1-2-40.ap-south-1.compute.internal:9092, ip-10-1-2-24.ap-south-1.compute.internal:9092**

23/05/31 16:53:50 INFO utils.Log4jControllerRegistration$: Registered kafka:type=kafka.Log4jController MBean

23/05/31 16:53:50 INFO producer.ProducerConfig: ProducerConfig values:

acks = 1

batch.size = 16384

bootstrap.servers = [ip-10-1-1-204.ap-south-1.compute.internal:9092, ]

buffer.memory = 33554432

client.dns.lookup = default

client.id = console-producer

compression.type = none

connections.max.idle.ms = 540000

delivery.timeout.ms = 120000

enable.idempotence = false

interceptor.classes = []

key.serializer = class org.apache.kafka.common.serialization.ByteArraySerializer

linger.ms = 1000

max.block.ms = 60000

max.in.flight.requests.per.connection = 5

max.request.size = 1048576

metadata.max.age.ms = 300000

metric.reporters = []

metrics.num.samples = 2

metrics.recording.level = INFO

metrics.sample.window.ms = 30000

partitioner.class = class org.apache.kafka.clients.producer.internals.DefaultPartitioner

receive.buffer.bytes = 32768

reconnect.backoff.max.ms = 1000

reconnect.backoff.ms = 50

request.timeout.ms = 1500

retries = 3

retry.backoff.ms = 100

sasl.client.callback.handler.class = null

sasl.jaas.config = null

sasl.kerberos.kinit.cmd = /usr/bin/kinit

sasl.kerberos.min.time.before.relogin = 60000

sasl.kerberos.service.name = null

sasl.kerberos.ticket.renew.jitter = 0.05

sasl.kerberos.ticket.renew.window.factor = 0.8

sasl.login.callback.handler.class = null

sasl.login.class = null

sasl.login.refresh.buffer.seconds = 300

sasl.login.refresh.min.period.seconds = 60

sasl.login.refresh.window.factor = 0.8

sasl.login.refresh.window.jitter = 0.05

sasl.mechanism = GSSAPI

security.protocol = PLAINTEXT

send.buffer.bytes = 102400

ssl.cipher.suites = null

ssl.enabled.protocols = [TLSv1.2, TLSv1.1, TLSv1]

ssl.endpoint.identification.algorithm = null

ssl.key.password = null

ssl.keymanager.algorithm = SunX509

ssl.keystore.location = null

ssl.keystore.password = null

ssl.keystore.type = JKS

ssl.protocol = TLS

ssl.provider = null

ssl.secure.random.implementation = null

ssl.trustmanager.algorithm = PKIX

ssl.truststore.location = null

ssl.truststore.password = null

ssl.truststore.type = JKS

transaction.timeout.ms = 60000

transactional.id = null

value.serializer = class org.apache.kafka.common.serialization.ByteArraySerializer

23/05/31 16:53:50 INFO utils.AppInfoParser: Kafka version : 2.1.0-cdh6.2.1

23/05/31 16:53:50 INFO utils.AppInfoParser: Kafka commitId : unknown

>

*From <*[*https://bigdata.cloudlabs.edureka.co/shell/*](https://bigdata.cloudlabs.edureka.co/shell/)*>*

[raghavendr48edu@ip-10-1-1-204 data]$ **vi kafka\_flume.conf**

[raghavendr48edu@ip-10-1-1-204 data]$ **cat kafka\_flume.conf**

fkgent.sources = source1

fkagent.channels = channel1 #source1 connecting to channel1

fkagent.sinks = sink1

fkagent.sources.source1.type = org.apache.flume.source.kafka.KafkaSource

fkagent.sources.source1.kafka.bootstrap.servers=ip-10-1-1-204.ap-south-1.compute.internal:9092

fkagent.sources.source1.kafka.topics=Flume\_Topic

fkagent.sources.source1.kafka.consumer.group.id=flume

fkagent.sources.source1.channels=channel1

fkagent.sources.source1.interceptors = i

fkagent.sources.source1.interceptors.i1.type = timestamp

fkagent.sources.source1.kafka.consumer.timeout.ms=100 #how often different files are created

fkagent.channels.channel1.type=memory #we can define as hdfs

fkagent.channels.channel1.capacity=10000

fkagent.channels.channel1.transactionCapacity=1000

fkagent.sinks.sink1.type=hdfs

fkagent.sinks.sink1.hdfs.path=hdfs://nameservice1/user/raghavendr48edu/flume\_kafka\_text

fkagent.sinks.sink1.hdfs.rollInterval=5 #every 5 sec file will be created

fkagent.sinks.sink1.hdfs.rollSize=0

fkagent.sinks.sink1.hdfs.rollCount=0

fkagent.sinks.sink1.hdfs.fileType=DataStream

fkagent.sinks.sink1.channel=channel1

[raghavendr48edu@ip-10-1-1-204 data]$

*From <*[*https://bigdata.cloudlabs.edureka.co/shell/*](https://bigdata.cloudlabs.edureka.co/shell/)*>*

[raghavendr48edu@ip-10-1-1-204 ~]$ **hdfs dfs -mkdir /user/raghavendr48edu/flume\_kafka\_test**

[raghavendr48edu@ip-10-1-1-204 ~]$ hdfs dfs -cat /user/raghavendr48edu/flume\_kafka\_test

[raghavendr48edu@ip-10-1-1-204 ~]$ **cd /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/bin**

[raghavendr48edu@ip-10-1-1-204 bin]$ ls

**avro-tools** **impala-collect-diagnostics** **kafka-preferred-replica-election** **oozie** **sqoop-import-all-tables**

**beeline** **impala-collect-minidumps** **kafka-producer-perf-test** **oozie-setup** **sqoop-job**

**bigtop-detect-javahome** **impalad** **kafka-reassign-partitions** **parquet-tools** **sqoop-list-databases**

**catalogd** **impala-shell** **kafka-replica-verification** **pig** **sqoop-list-tables**

**cli\_mt** **kafka-acls** **kafka-run-class** **pyspark** **sqoop-merge**

**cli\_st** **kafka-broker-api-versions** **kafka-sentry** **sentry** **sqoop-metastore**

**flume-ng** **kafka-configs** **kafka-server-start** **solrctl** **sqoop-version**

**hadoop** **kafka-console-consumer** **kafka-server-stop** **spark-shell** **statestored**

**hadoop-fuse-dfs** **kafka-console-producer** **kafka-streams-application-reset** **spark-submit** **yarn**

**hbase** **kafka-consumer-groups** **kafka-topics** **sqoop** **zookeeper-client**

**hbase-indexer** **kafka-consumer-perf-test** **kafka-verifiable-consumer** **sqoop-codegen** **zookeeper-security-migration**

**hbase-indexer-sentry** **kafka-delegation-tokens** **kafka-verifiable-producer** **sqoop-create-hive-table** **zookeeper-server**

**hcat** **kafka-delete-records** **kite-dataset** **sqoop-eval** **zookeeper-server-cleanup**

**hdfs** **kafka-dump-log** **kudu** **sqoop-export** **zookeeper-server-initialize**

**hive** **kafka-log-dirs** **load\_gen** **sqoop-help**

**hiveserver2** **kafka-mirror-maker** **mapred** **sqoop-import**

[raghavendr48edu@ip-10-1-1-204 bin]$

*From <*[*https://bigdata.cloudlabs.edureka.co/shell/*](https://bigdata.cloudlabs.edureka.co/shell/)*>*

/flume-ng agent --conf conf

(it uses configuration file

**./flume-ng agent --conf conf --conf-file /home/raghavendr48edu/data/kafka\_flume.conf --name fkagent -Dflume.root.logger=INFO.console**

(path of conf file)

[raghavendr48edu@ip-10-1-1-204 bin]$**./flume-ng agent --conf conf --conf-file /home/raghavendr48edu/data/kafka\_flume.conf --name fkagent -Dflume.root.logger=INFO.console**

23/05/31 06:21:35 WARN consumer.ConsumerConfig: The configuration 'timeout.ms' was supplied but isn't a known config.

23/05/31 06:21:35 INFO utils.AppInfoParser: Kafka version : 2.1.0-cdh6.2.1

23/05/31 06:21:35 INFO utils.AppInfoParser: Kafka commitId : null

23/05/31 06:21:35 INFO kafka.KafkaSource: Kafka source source1 started.

23/05/31 06:21:35 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SOURCE, name: source1: Successfully registered new MBean.

23/05/31 06:21:35 INFO instrumentation.MonitoredCounterGroup: Component type: SOURCE, name: source1 started

23/05/31 06:21:35 INFO clients.Metadata: Cluster ID: i6LyuYbbSMiPQc6ARDUuCg

23/05/31 06:21:35 INFO internals.AbstractCoordinator: [Consumer clientId=consumer-1, groupId=flume] Discovered group coordinator ip-10-1-2-40.ap-south-1.comp

ute.internal:9092 (id: 2147483460 rack: null)

23/05/31 06:21:35 INFO internals.ConsumerCoordinator: [Consumer clientId=consumer-1, groupId=flume] Revoking previously assigned partitions []

23/05/31 06:21:35 INFO internals.AbstractCoordinator: [Consumer clientId=consumer-1, groupId=flume] (Re-)joining group

23/05/31 06:21:40 INFO internals.AbstractCoordinator: [Consumer clientId=consumer-1, groupId=flume] Successfully joined group with generation 1

23/05/31 06:21:40 INFO internals.ConsumerCoordinator: [Consumer clientId=consumer-1, groupId=flume] Setting newly assigned partitions [Flume\_Topic-0]

23/05/31 06:21:40 INFO kafka.SourceRebalanceListener: topic Flume\_Topic - partition 0 assigned.

23/05/31 06:21:40 INFO internals.Fetcher: [Consumer clientId=consumer-1, groupId=flume] Resetting offset for partition Flume\_Topic-0 to offset 0.

[raghavendr48edu@ip-10-1-1-204 ~]$ **kafka-console-producer --topic Flume\_Topic --broker-list ip-10-1-1-204.ap-south-1.compute.internal:9092, ip-10-1-2-40.ap-south-1.compute.internal:9092, ip-10-1-2-24.ap-south-1.compute.internal:9092**

23/05/31 06:24:19 INFO utils.Log4jControllerRegistration$: Registered kafka:type=kafka.Log4jController MBean

23/05/31 06:24:19 INFO producer.ProducerConfig: ProducerConfig values:

acks = 1

batch.size = 16384

bootstrap.servers = [ip-10-1-1-204.ap-south-1.compute.internal:9092, ]

buffer.memory = 33554432

client.dns.lookup = default

client.id = console-producer

compression.type = none

connections.max.idle.ms = 540000

delivery.timeout.ms = 120000

enable.idempotence = false

interceptor.classes = []

key.serializer = class org.apache.kafka.common.serialization.ByteArraySerializer

linger.ms = 1000

max.block.ms = 60000

max.in.flight.requests.per.connection = 5

max.request.size = 1048576

metadata.max.age.ms = 300000

metric.reporters = []

metrics.num.samples = 2

metrics.recording.level = INFO

metrics.sample.window.ms = 30000

partitioner.class = class org.apache.kafka.clients.producer.internals.DefaultPartitioner

receive.buffer.bytes = 32768

reconnect.backoff.max.ms = 1000

reconnect.backoff.ms = 50

request.timeout.ms = 1500

retries = 3

retry.backoff.ms = 100

sasl.client.callback.handler.class = null

sasl.jaas.config = null

sasl.kerberos.kinit.cmd = /usr/bin/kinit

sasl.kerberos.min.time.before.relogin = 60000

sasl.kerberos.service.name = null

sasl.kerberos.ticket.renew.jitter = 0.05

sasl.kerberos.ticket.renew.window.factor = 0.8

sasl.login.callback.handler.class = null

sasl.login.class = null

sasl.login.refresh.buffer.seconds = 300

sasl.login.refresh.min.period.seconds = 60

sasl.login.refresh.window.factor = 0.8

sasl.login.refresh.window.jitter = 0.05

sasl.mechanism = GSSAPI

security.protocol = PLAINTEXT

send.buffer.bytes = 102400

ssl.cipher.suites = null

ssl.enabled.protocols = [TLSv1.2, TLSv1.1, TLSv1]

ssl.endpoint.identification.algorithm = null

ssl.key.password = null

ssl.keymanager.algorithm = SunX509

ssl.keystore.location = null

ssl.keystore.password = null

ssl.keystore.type = JKS

ssl.protocol = TLS

ssl.provider = null

ssl.secure.random.implementation = null

ssl.trustmanager.algorithm = PKIX

ssl.truststore.location = null

ssl.truststore.password = null

ssl.truststore.type = JKS

transaction.timeout.ms = 60000

transactional.id = null

value.serializer = class org.apache.kafka.common.serialization.ByteArraySerializer

23/05/31 06:24:19 INFO utils.AppInfoParser: Kafka version : 2.1.0-cdh6.2.1

23/05/31 06:24:19 INFO utils.AppInfoParser: Kafka commitId : unknown

>hello everyone

23/05/31 06:24:30 INFO clients.Metadata: Cluster ID: i6LyuYbbSMiPQc6ARDUuCg

>raghav is

>practising

>kafka

>sqoop

>hadoop

>

*From <*[*https://bigdata.cloudlabs.edureka.co/shell/*](https://bigdata.cloudlabs.edureka.co/shell/)*>*

23/05/31 06:21:35 WARN consumer.ConsumerConfig: The configuration 'timeout.ms' was supplied but isn't a known config.

23/05/31 06:21:35 INFO utils.AppInfoParser: Kafka version : 2.1.0-cdh6.2.1

23/05/31 06:21:35 INFO utils.AppInfoParser: Kafka commitId : null

23/05/31 06:21:35 INFO kafka.KafkaSource: Kafka source source1 started.

23/05/31 06:21:35 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SOURCE, name: source1: Successfully registered new MBean.

23/05/31 06:21:35 INFO instrumentation.MonitoredCounterGroup: Component type: SOURCE, name: source1 started

23/05/31 06:21:35 INFO clients.Metadata: Cluster ID: i6LyuYbbSMiPQc6ARDUuCg

23/05/31 06:21:35 INFO internals.AbstractCoordinator: [Consumer clientId=consumer-1, groupId=flume] Discovered group coordinator ip-10-1-2-40.ap-south-1.comp

ute.internal:9092 (id: 2147483460 rack: null)

23/05/31 06:21:35 INFO internals.ConsumerCoordinator: [Consumer clientId=consumer-1, groupId=flume] Revoking previously assigned partitions []

23/05/31 06:21:35 INFO internals.AbstractCoordinator: [Consumer clientId=consumer-1, groupId=flume] (Re-)joining group

23/05/31 06:21:40 INFO internals.AbstractCoordinator: [Consumer clientId=consumer-1, groupId=flume] Successfully joined group with generation 1

23/05/31 06:21:40 INFO internals.ConsumerCoordinator: [Consumer clientId=consumer-1, groupId=flume] Setting newly assigned partitions [Flume\_Topic-0]

23/05/31 06:21:40 INFO kafka.SourceRebalanceListener: topic Flume\_Topic - partition 0 assigned.

23/05/31 06:21:40 INFO internals.Fetcher: [Consumer clientId=consumer-1, groupId=flume] Resetting offset for partition Flume\_Topic-0 to offset 0.

23/05/31 06:24:37 INFO hdfs.HDFSDataStream: Serializer = TEXT, UseRawLocalFileSystem = false

23/05/31 06:24:37 INFO hdfs.BucketWriter: Creating hdfs://nameservice1/user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514277124.tmp

23/05/31 06:24:43 INFO hdfs.HDFSEventSink: Writer callback called.

23/05/31 06:24:43 INFO hdfs.BucketWriter: Closing hdfs://nameservice1/user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514277124.tmp

23/05/31 06:24:43 INFO hdfs.BucketWriter: Renaming hdfs://nameservice1/user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514277124.tmp to hdfs://nameservic

e1/user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514277124

23/05/31 06:24:45 INFO hdfs.HDFSDataStream: Serializer = TEXT, UseRawLocalFileSystem = false

23/05/31 06:24:45 INFO hdfs.BucketWriter: Creating hdfs://nameservice1/user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514285469.tmp

23/05/31 06:24:50 INFO hdfs.HDFSEventSink: Writer callback called.

23/05/31 06:24:50 INFO hdfs.BucketWriter: Closing hdfs://nameservice1/user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514285469.tmp

23/05/31 06:24:50 INFO hdfs.BucketWriter: Renaming hdfs://nameservice1/user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514285469.tmp to hdfs://nameservic

e1/user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514285469

23/05/31 06:24:55 INFO hdfs.HDFSDataStream: Serializer = TEXT, UseRawLocalFileSystem = false

23/05/31 06:24:55 INFO hdfs.BucketWriter: Creating hdfs://nameservice1/user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514295477.tmp

23/05/31 06:25:00 INFO hdfs.HDFSEventSink: Writer callback called.

23/05/31 06:25:00 INFO hdfs.BucketWriter: Closing hdfs://nameservice1/user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514295477.tmp

23/05/31 06:25:00 INFO hdfs.BucketWriter: Renaming hdfs://nameservice1/user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514295477.tmp to hdfs://nameservic

e1/user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514295477

*From <*[*https://bigdata.cloudlabs.edureka.co/shell/*](https://bigdata.cloudlabs.edureka.co/shell/)*>*

[raghavendr48edu@ip-10-1-1-204 ~]$ **hdfs dfs -ls /user/raghavendr48edu/flume\_kafka\_test**

Found 3 items

-rw-r--r-- 3 raghavendr48edu raghavendr48edu 25 2023-05-31 06:24 /user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514277124

-rw-r--r-- 3 raghavendr48edu raghavendr48edu 17 2023-05-31 06:24 /user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514285469

-rw-r--r-- 3 raghavendr48edu raghavendr48edu 13 2023-05-31 06:25 /user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514295477

[raghavendr48edu@ip-10-1-1-204 ~]$

*From <*[*https://bigdata.cloudlabs.edureka.co/shell/*](https://bigdata.cloudlabs.edureka.co/shell/)*>*

3 files created because there is some difference in milliseconds I provided as my conf parameter and it would create file after those milllisec

[raghavendr48edu@ip-10-1-1-204 ~]$ hdfs dfs -cat /user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514277124

hello everyone

raghav is

*From <*[*https://bigdata.cloudlabs.edureka.co/shell/*](https://bigdata.cloudlabs.edureka.co/shell/)*>*

[raghavendr48edu@ip-10-1-1-204 ~]$ hdfs dfs -cat /user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514285469

practising

kafka

*From <*[*https://bigdata.cloudlabs.edureka.co/shell/*](https://bigdata.cloudlabs.edureka.co/shell/)*>*

[raghavendr48edu@ip-10-1-1-204 ~]$ hdfs dfs -cat /user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514295477

sqoop

hadoop

*From <*[*https://bigdata.cloudlabs.edureka.co/shell/*](https://bigdata.cloudlabs.edureka.co/shell/)*>*

23/05/31 06:24:30 INFO clients.Metadata: Cluster ID: i6LyuYbbSMiPQc6ARDUuCg

>raghav is

>practising

>kafka

>sqoop

>hadoop

>Now

>everyone

>knows

>how

>flume

>is

>used in

>big data world

*From <*[*https://bigdata.cloudlabs.edureka.co/shell/*](https://bigdata.cloudlabs.edureka.co/shell/)*>*

e1/user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514295477

23/05/31 06:33:22 INFO hdfs.HDFSDataStream: Serializer = TEXT, UseRawLocalFileSystem = false

23/05/31 06:33:22 INFO hdfs.BucketWriter: Creating hdfs://nameservice1/user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514802582.tmp

23/05/31 06:33:27 INFO hdfs.HDFSEventSink: Writer callback called.

23/05/31 06:33:27 INFO hdfs.BucketWriter: Closing hdfs://nameservice1/user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514802582.tmp

23/05/31 06:33:27 INFO hdfs.BucketWriter: Renaming hdfs://nameservice1/user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514802582.tmp to hdfs://nameservic

e1/user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514802582

23/05/31 06:33:32 INFO hdfs.HDFSDataStream: Serializer = TEXT, UseRawLocalFileSystem = false

23/05/31 06:33:32 INFO hdfs.BucketWriter: Creating hdfs://nameservice1/user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514812588.tmp

23/05/31 06:33:37 INFO hdfs.HDFSEventSink: Writer callback called.

23/05/31 06:33:37 INFO hdfs.BucketWriter: Closing hdfs://nameservice1/user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514812588.tmp

23/05/31 06:33:37 INFO hdfs.BucketWriter: Renaming hdfs://nameservice1/user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514812588.tmp to hdfs://nameservic

e1/user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514812588

*From <*[*https://bigdata.cloudlabs.edureka.co/shell/*](https://bigdata.cloudlabs.edureka.co/shell/)*>*

[raghavendr48edu@ip-10-1-1-204 ~]$ hdfs dfs -ls /user/raghavendr48edu/flume\_kafka\_test

Found 5 items

-rw-r--r-- 3 raghavendr48edu raghavendr48edu 25 2023-05-31 06:24 /user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514277124

-rw-r--r-- 3 raghavendr48edu raghavendr48edu 17 2023-05-31 06:24 /user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514285469

-rw-r--r-- 3 raghavendr48edu raghavendr48edu 13 2023-05-31 06:25 /user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514295477

-rw-r--r-- 3 raghavendr48edu raghavendr48edu 33 2023-05-31 06:33 /user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514802582

-rw-r--r-- 3 raghavendr48edu raghavendr48edu 23 2023-05-31 06:33 /user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514812588

[raghavendr48edu@ip-10-1-1-204 ~]$

*From <*[*https://bigdata.cloudlabs.edureka.co/shell/*](https://bigdata.cloudlabs.edureka.co/shell/)*>*

[raghavendr48edu@ip-10-1-1-204 ~]$ hdfs dfs -cat /user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514802582

Now

everyone

knows

how

flume

is

*From <*[*https://bigdata.cloudlabs.edureka.co/shell/*](https://bigdata.cloudlabs.edureka.co/shell/)*>*

[raghavendr48edu@ip-10-1-1-204 ~]$ hdfs dfs -cat /user/raghavendr48edu/flume\_kafka\_test/FlumeData.1685514812588

used in

big data world

*From <*[*https://bigdata.cloudlabs.edureka.co/shell/*](https://bigdata.cloudlabs.edureka.co/shell/)*>*