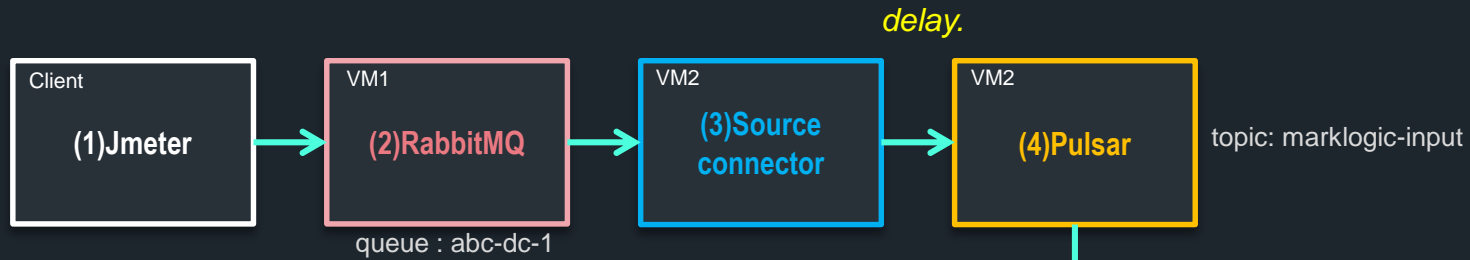
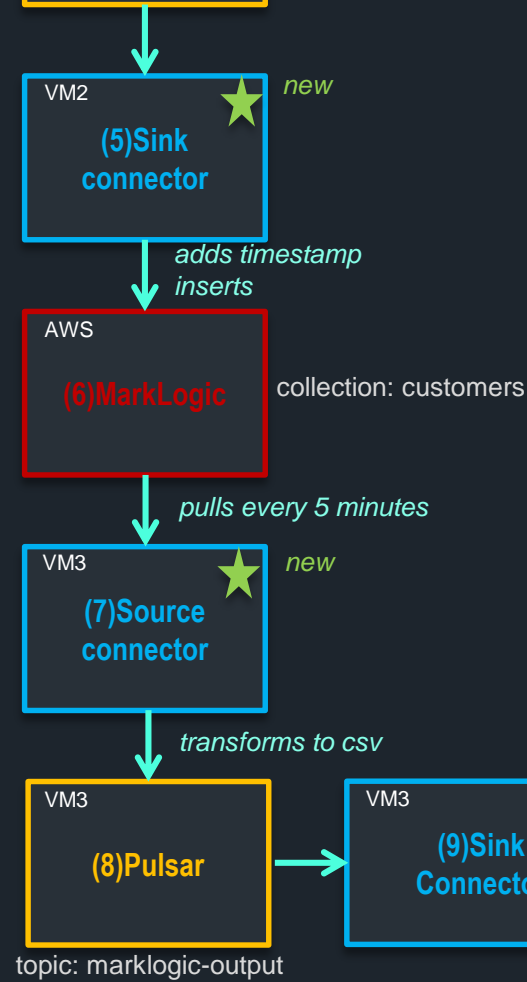


Data Pipeline - 1



1. Client generates JSON messages into RabbitMQ (VM1)
2. Store the messages in MarkLogic (AWS) in real time
3. Replicate the data to a target RabbitMQ with a maximum delay of 5 minutes.



Components

Component	Description
(1)Jmeter	The client that produces JSON data for the RabbitMQ through exchange <i>ABC-DC1</i> bound to queue <i>abc-dc-1</i> with routing key <i>abc.dc.1</i>
(2)RabbitMQ	The first component of the data pipeline which accepts JSON messages from the Jmeter client. The topic <i>abc-dc-1</i> is transient.
(3)Source Connector	The built-in RabbitMQ Source connector that pushes JSON messages from <i>abc-dc-1</i> to pulsar topic <i>marklogic-input</i> .
(4)Pulsar	Apache pulsar with topic <i>marklogic-input</i>
(5)Sink Connector	The newly built MarkLogic Sink Connector that pushes JSON messages from pulsar topic <i>marklogic-input</i> into MarkLogic. When inserting into MarkLogic, a timestamp attribute is added to the data using a transform function. Authentication, URI format, collections, permissions etc are configured in the Sink Connector configuration file. If there is no transform function, then the JSON messages will be inserted as it is.
(6)MarkLogic	The MarkLogic database in the pipeline that stores JSON documents pulled from Pulsar topic <i>marklogic-input</i> .
(7)Source Connector	The newly built MarkLogic Source Connector that pulls JSON messages from MarkLogic using a configured query. The pull from MarkLogic is triggered through a CRON job that triggers based on a configurable CRON expression. When pulled from MarkLogic the JSON document is converted to CSV format using a transform function. The transform function can be used to reshape the document to any format. The connector writes to Pulsar topic <i>marklogic-output</i> . If the transform function is not configured, then JSON messages from MarkLogic will be pushed to RabbitMQ as-is.
(8)Pulsar	Apache Pulsar with topic <i>marklogic-output</i> that stores the CSV converted data from MarkLogic.
(9)Sink Connector	The built-in RabbitMQ Sink Connector that CSV messages from topic <i>marklogic-output</i> into RabbitMQ exchange <i>ABC-DC2</i> bound to queue <i>abc-dc-2</i> with routing key <i>abc.dc.2</i>
(10)RabbitMQ	The final component of the data pipeline that stores the CSV messages from <i>marklogic-output</i> .



Connector configurations

Marklogic Source Connector Configuration

```
{
  "configs": {
    "mlPathToCertFile": "",
    "mlSecurityContext": "BASIC",
    "mlSSLMutualAuth": "false",
    "dmsdkTransformParams": "",
    "mlConnectionHost": "<mymarklogicendpoint>",
    "dmsdkTransform": "CSVTransform",
    "mlUserName": "dh-admin-1",
    "dmsdkSourceQuery":
      "cts.andQuery([cts.collectionQuery('customers'),cts.pathRangeQuery('//insertTS', '<', fn.currentDateTime()),cts.pathRangeQuery('//insertTS', '>=', fn.currentDateTime().subtract('PT5M'))])",
    "dmsdkIsSourceQuerySerialized": "false",
    "mlHostNameVerifier": "ANY",
    "mlPassword": "MarkLogic*001",
    "mlSimpleSSL": "true",
    "dmsdkBatchSize": 100,
    "mlExternalName": "",
    "mlTlsVersion": "TLSv1.2",
    "mlConnectionPort": 8010,
    "dmsdkThreadCount": 10,
    "mlDatabase": "data-hub-STAGING",
    "mlPasswordForCertFile": "",
    "mlConnectionType": "GATEWAY"
  },
  "batchSourceConfig": {
    "discoveryTriggererClassName":
      "com.marklogic.pulsar.config.CronTriggerer",
    "discoveryTriggererConfig": {
      "__CRON__": "0 0/5 * * * ?"
    }
  }
}
```

RabbitMQ Source Connector Configuration

```
{
  "configs": {
    "prefetchCount": "0",
    "requestedChannelMax": "0",
    "prefetchGlobal": "false",
    "handshakeTimeout": "10000",
    "virtualHost": "/",
    "requestedFrameMax": "0",
    "password": "admin",
    "queueName": "abc-dc-1",
    "port": "5672",
    "host": "192.168.0.35",
    "requestedHeartbeat": "60",
    "connectionName": "pulsar source",
    "connectionTimeout": "60000",
    "username": "admin"
  }
}
```

RabbitMQ Sink Connector Configuration

```
{
  "configs": {
    "requestedChannelMax": "0",
    "handshakeTimeout": "10000",
    "virtualHost": "/",
    "requestedFrameMax": "0",
    "password": "admin",
    "queueName": "abc-dc-2",
    "port": "5672",
    "host": "192.168.0.35",
    "requestedHeartbeat": "60",
    "exchangeName": "ABC-DC2",
    "connectionName": "test-connection",
    "connectionTimeout": "60000",
    "routingKey": "abc.dc.2",
    "username": "admin"
  }
}
```



Connector configurations

Marklogic Sink Connector Configuration

```
{
  "configs": {
    "mlPathToCertFile": "",
    "mlSecurityContext": "BASIC",
    "mlSSLMutualAuth": "false",
    "mlConnectionHost": "mymarklogicendpoint ",
    "mlDocumentFormat": "JSON",
    "mlAddTopicAsCollections": "true",
    "mlPassword": "MarkLogic*001",
    "dmsdkBatchSize": 1,
    "mlExternalName": "",
    "mlDocumentMimeType": "",
    "mlIdStrategyPath": "/CustomerInfo/key",
    "mlConnectionPort": 8010,
    "mlDatabase": "data-hub-STAGING",
    "mlConnectionType": "GATEWAY",
    "dmsdkTransformParams": "",
    "mlIdStrategyForURI": "JSONPATH",
    "dmsdkTransform": "insertTSTransform",
    "mlUserName": "dh-admin-1",
    "mlHostNameVerifier": "ANY",
    "mlDocumentPermissions": "rest-reader,read,rest-writer,update",
    "mlSimpleSSL": "true",
    "mlDocumentURIPrefix": "/pulsar-data/customers/",
    "mlTlsVersion": "TLSv1.2",
    "mlDocumentURISuffix": ".json",
    "dmsdkThreadCount": 10,
    "mlDocumentCollections": "customers",
    "mlPasswordForCertFile": ""
  }
}
```

