

SMOKE TEST

StreamSets Data Collector

Date Prepared: May 2020

Document Information

| | | | |
|-----------------------|---|---------------------|----------|
| Project Name | EPIC Accelerator Deployment & Integration Services | | |
| Project Owner | | Document Version No | 0.1 |
| Quality Review Method | | | |
| Prepared By | Prasad Adireddi | Preparation Date | May 2020 |
| Reviewed By | | Review Date | |

Table of Contents

1 LOGIN TO STREAMSETS DATA COLLECTOR4

2 TESTING LOCAL FS TO HADOOP FS PIPELINE EXAMPLE6

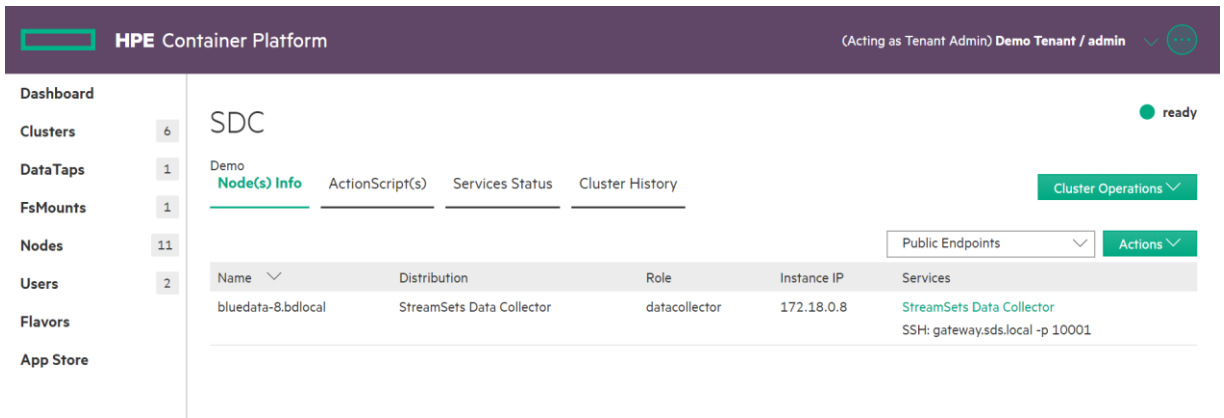
Table of Tables

NO TABLE OF FIGURES ENTRIES FOUND.

1 LOGIN TO STREAMSETS DATA COLLECTOR

Login to Jupyter Notebook, using the following steps:

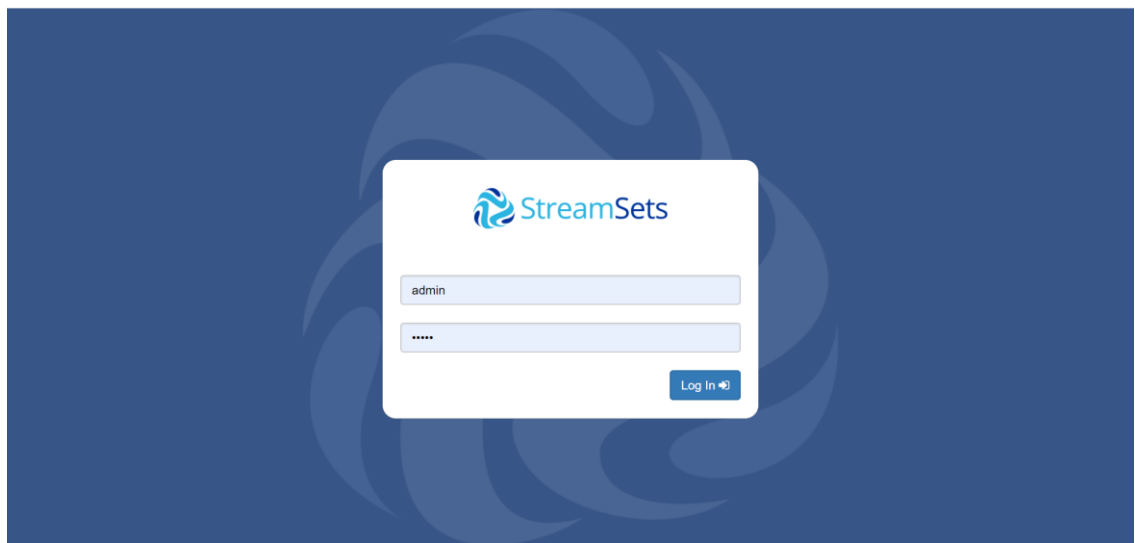
1. From the Cluster page under **SERVICES** column, click on **StreamSets Data Collector**



The screenshot shows the HPE Container Platform interface. The left sidebar contains navigation links: Dashboard, Clusters (6), DataTaps (1), FsMounts (1), Nodes (11), Users (2), Flavors, and App Store. The main content area is titled 'SDC' and shows a table of services. The table has columns: Name, Distribution, Role, Instance IP, and Services. The 'StreamSets Data Collector' service is listed under the 'bluedata-8.bdllocal' distribution. The status is 'ready'.

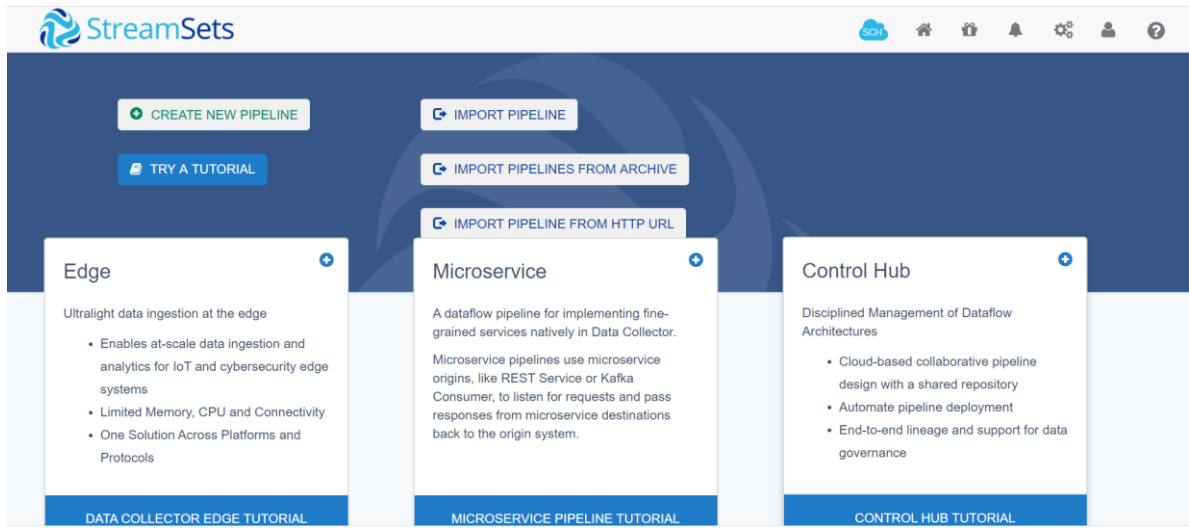
| Name | Distribution | Role | Instance IP | Services |
|---------------------|---------------------------|---------------|-------------|--|
| bluedata-8.bdllocal | StreamSets Data Collector | datacollector | 172.18.0.8 | StreamSets Data Collector SSH: gateway.sds.local -p 10001 |

2. It will open-up a new tab with SDC login page



3. Enter the username and password (admin/admin) and click on **Log in**
4. You will get SDC Home page with some tutorial options

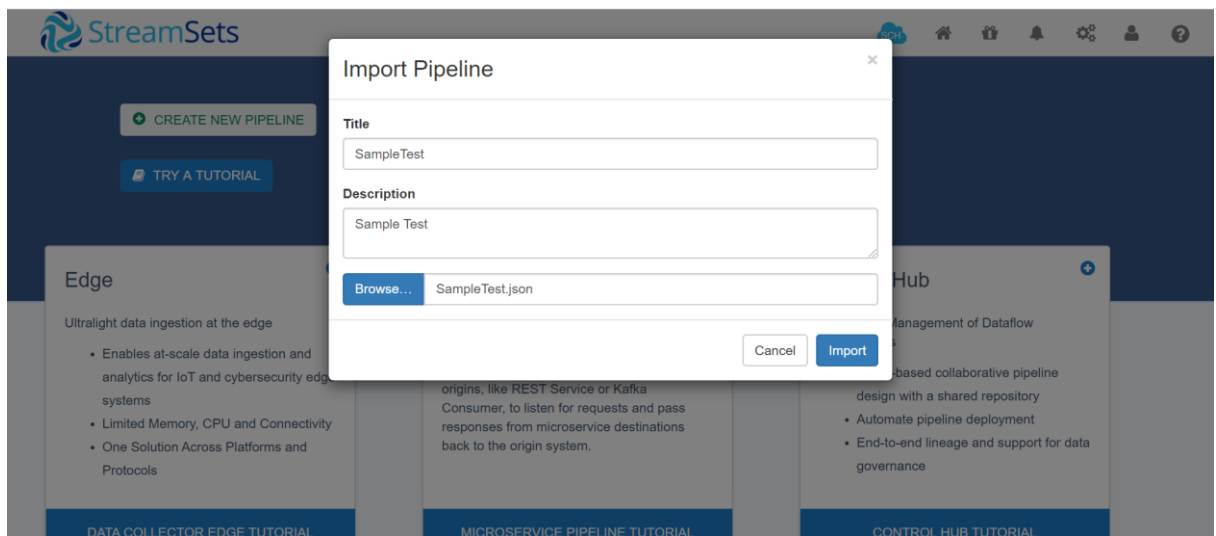
Smoke Test Document StreamSets Data Collector



2 TESTING LOCAL FS TO HADOOP FS PIPELINE EXAMPLE

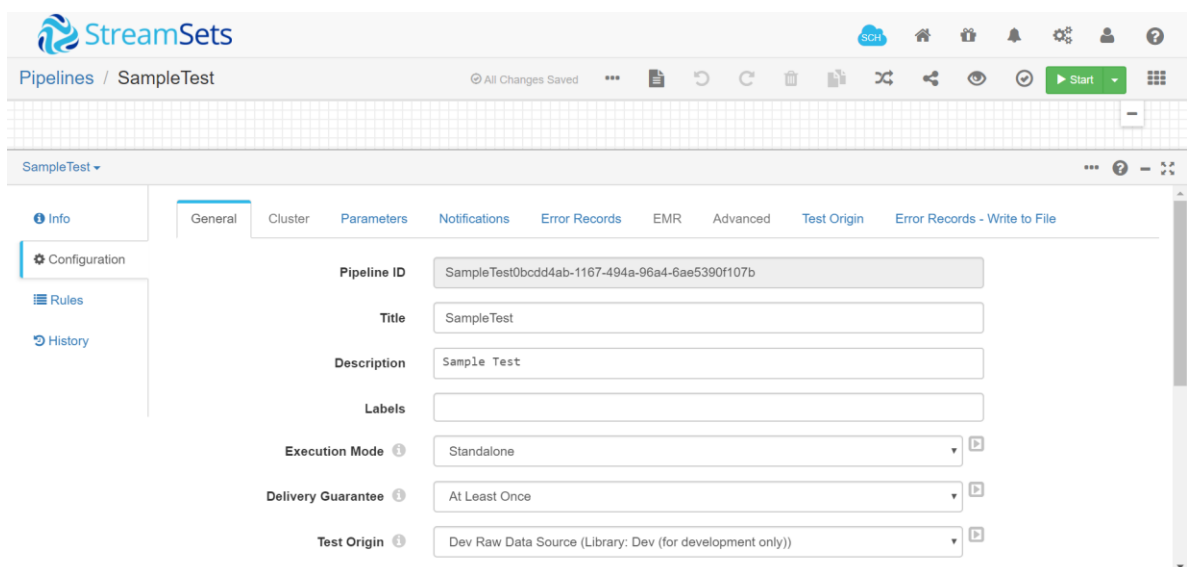
Pipeline streams data from origin local filesystem to remote Hadoop filesystem:

1. Copy the JSON data from APPENDIX A and save it as .json file.
2. From the SDC Home page, click **IMPORT PIPELINE** and upload the json file



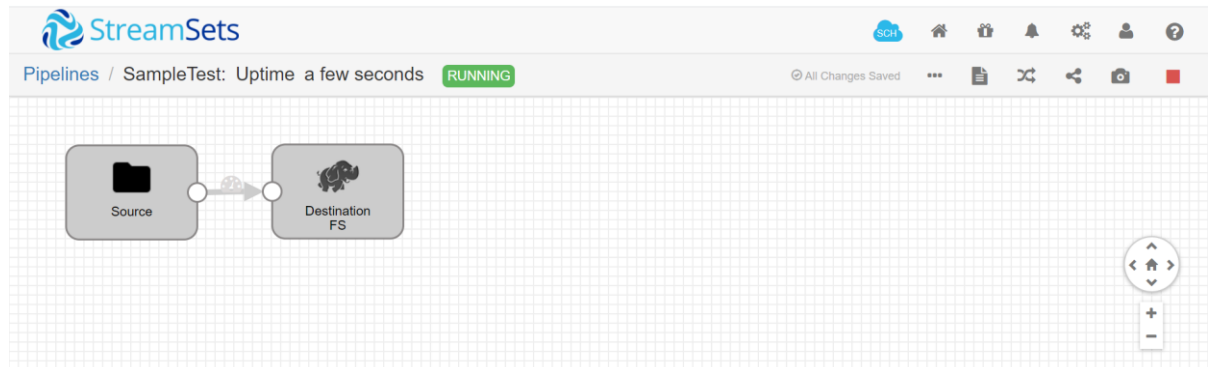
2.1 Validate Imported Pipeline

Click on validate button to validate the pipeline source and destination configurations.



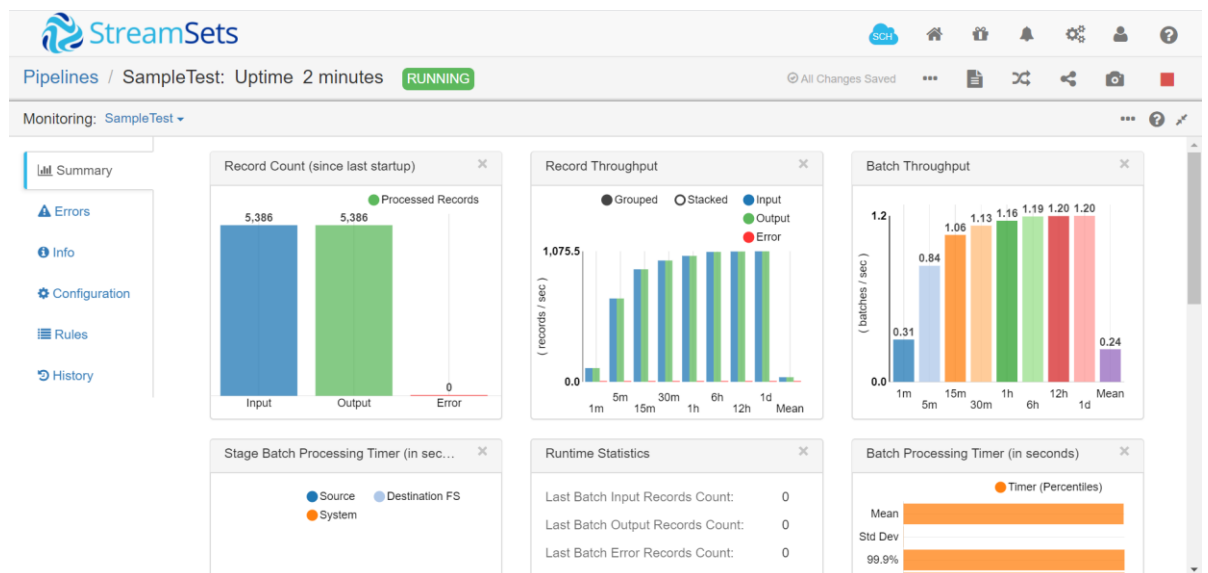
2.2 Running Pipeline

Click on the 'Start' to execute the pipeline



2.3 Pipeline execution report

Pipeline streaming summary can be validated at monitoring window



APPENDIX A: Sample JSON file for StreamSets data pipeline

```
{
  "pipelineConfig" : {
    "schemaVersion" : 6,
    "version" : 17,
    "pipelineId" : "SampleTest6929ae91-0b30-4a8e-b111-1d64c778ce09",
    "title" : "SampleTest",
    "description" : "test pipeline",
    "uuid" : "29e035a7-2da3-4506-9980-1e7e66f72031",
    "configuration" : [ {
      "name" : "executionMode",
      "value" : "STANDALONE"
    }, {
      "name" : "edgeHttpUrl",
      "value" : "http://localhost:18633"
    }, {
      "name" : "deliveryGuarantee",
      "value" : "AT_LEAST_ONCE"
    }, {
      "name" : "testOriginStage",
      "value" : "streamsets-datacollector-dev-lib::com_streamsets_pipeline_stage_devtest_rawdata_RawDataDSource::3"
    }, {
      "name" : "startEventStage",
      "value" : "streamsets-datacollector-basic-lib::com_streamsets_pipeline_stage_destination_devnull_ToErrorNullDTarget::1"
    }, {
      "name" : "stopEventStage",
      "value" : "streamsets-datacollector-basic-lib::com_streamsets_pipeline_stage_destination_devnull_ToErrorNullDTarget::1"
    }, {
      "name" : "shouldRetry",
      "value" : true
    }, {
      "name" : "triggerInterval",
      "value" : 2000
    }, {
      "name" : "retryAttempts",
      "value" : -1
    }, {
      "name" : "ludicrousMode",
      "value" : false
    }, {
      "name" : "ludicrousModeInputCount",
      "value" : false
    }, {
      "name" : "advancedErrorHandling",
      "value" : false
    }, {
      "name" : "notifyOnStates",
      "value" : [ "RUN_ERROR", "STOPPED", "FINISHED" ]
    }
  ]
}
```



```

    }, {
      "name" : "emailIDs",
      "value" : [ ]
    }, {
      "name" : "constants",
      "value" : [ ]
    }, {
      "name" : "badRecordsHandling",
      "value" : "streamsets-datacollector-basic-
lib::com_streamsets_pipeline_stage_destination_recordstolocalfilesystem_ToErrorLocalFSDTarget::1"
    }, {
      "name" : "errorRecordPolicy",
      "value" : "ORIGINAL_RECORD"
    }, {
      "name" : "statsAggregatorStage",
      "value" : "streamsets-datacollector-basic-
lib::com_streamsets_pipeline_stage_destination_devnull_StatsDpmDirectlyDTarget::1"
    }, {
      "name" : "workerCount",
      "value" : 0
    }, {
      "name" : "clusterSlaveMemory",
      "value" : 2048
    }, {
      "name" : "clusterSlaveJavaOpts",
      "value" : "-XX:+UseConcMarkSweepGC -XX:+UseParNewGC -Dlog4j.debug"
    }, {
      "name" : "clusterLauncherEnv",
      "value" : [ ]
    }, {
      "name" : "mesosDispatcherURL",
      "value" : null
    }, {
      "name" : "logLevel",
      "value" : "INFO"
    }, {
      "name" : "hdfsS3ConfDir",
      "value" : null
    }, {
      "name" : "rateLimit",
      "value" : 0
    }, {
      "name" : "maxRunners",
      "value" : 0
    }, {
      "name" : "shouldCreateFailureSnapshot",
      "value" : true
    }, {
      "name" : "runnerIdleTime",
      "value" : 60
    }, {
      "name" : "webhookConfigs",
      "value" : [ ]
    }
  ]
}

```

```

    }, {
      "name" : "sparkConfigs",
      "value" : [ ]
    }, {
      "name" : "preprocessScript",
      "value" : "/*\n
The following script define a method\n
that increments an
integer by 1 \n
and registers it as a UDF with \n
the SparkSession, which can be
accessed\n
using the variable named \"spark\":\n
def inc(i: Integer): Integer =
{\n
  i + 1\n
}\n
spark.udf.register (\"inc\", inc _)\n
*/"
    }, {
      "name" : "clusterConfig.clusterType",
      "value" : "LOCAL"
    }, {
      "name" : "clusterConfig.sparkMasterUrl",
      "value" : "local[*]"
    }, {
      "name" : "clusterConfig.deployMode",
      "value" : "CLIENT"
    }, {
      "name" : "clusterConfig.hadoopUserName",
      "value" : null
    }, {
      "name" : "clusterConfig.sparkAppName",
      "value" : "${pipeline:title()}"
    }, {
      "name" : "clusterConfig.stagingDir",
      "value" : "/streamsets"
    }, {
      "name" : "clusterConfig.useYarnKerberosKeytab",
      "value" : false
    }, {
      "name" : "clusterConfig.yarnKerberosKeytabSource",
      "value" : "PROPERTIES_FILE"
    }, {
      "name" : "clusterConfig.yarnKerberosKeytab",
      "value" : null
    }, {
      "name" : "clusterConfig.yarnKerberosPrincipal",
      "value" : "name@DOMAIN"
    }, {
      "name" : "clusterConfig.callbackUrl",
      "value" : null
    }, {
      "name" : "databricksConfig.baseUrl",
      "value" : null
    }, {
      "name" : "databricksConfig.credentialType",
      "value" : null
    }, {
      "name" : "databricksConfig.username",
      "value" : ""
    }, {
      "name" : "databricksConfig.password",
      "value" : ""
    }, {

```

```

    "name" : "databricksConfig.token",
    "value" : ""
  }, {
    "name" : "databricksConfig.provisionNewCluster",
    "value" : true
  }, {
    "name" : "databricksConfig.clusterId",
    "value" : null
  }, {
    "name" : "databricksConfig.clusterConfig",
    "value" : "{\n      \"num_workers\": 8,\n      \"spark_version\": \"5.3.x-
scala2.11\", \n      \"node_type_id\": \"i3.xlarge\"\n}"
  }, {
    "name" : "databricksConfig.terminateCluster",
    "value" : false
  }, {
    "name" : "livyConfig.baseUrl",
    "value" : "https://localhost:30443/gateway/default/livy/v1/"
  }, {
    "name" : "livyConfig.username",
    "value" : ""
  }, {
    "name" : "livyConfig.password",
    "value" : ""
  }, {
    "name" : "amazonEMRConfig.userRegion",
    "value" : null
  }, {
    "name" : "amazonEMRConfig.userRegionCustom",
    "value" : null
  }, {
    "name" : "amazonEMRConfig.accessKey",
    "value" : ""
  }, {
    "name" : "amazonEMRConfig.secretKey",
    "value" : ""
  }, {
    "name" : "amazonEMRConfig.s3StagingUri",
    "value" : null
  }, {
    "name" : "amazonEMRConfig.provisionNewCluster",
    "value" : false
  }, {
    "name" : "amazonEMRConfig.clusterId",
    "value" : null
  }, {
    "name" : "amazonEMRConfig.clusterPrefix",
    "value" : null
  }, {
    "name" : "amazonEMRConfig.terminateCluster",
    "value" : false
  }, {
    "name" : "amazonEMRConfig.loggingEnabled",
    "value" : true
  }, {

```

```

    "name" : "amazonEMRConfig.s3LogUri",
    "value" : null
  }, {
    "name" : "amazonEMRConfig.enableEMRDebugging",
    "value" : true
  }, {
    "name" : "amazonEMRConfig.serviceRole",
    "value" : "EMR_DefaultRole"
  }, {
    "name" : "amazonEMRConfig.jobFlowRole",
    "value" : "EMR_EC2_DefaultRole"
  }, {
    "name" : "amazonEMRConfig.visibleToAllUsers",
    "value" : true
  }, {
    "name" : "amazonEMRConfig.ec2SubnetId",
    "value" : null
  }, {
    "name" : "amazonEMRConfig.masterSecurityGroup",
    "value" : null
  }, {
    "name" : "amazonEMRConfig.slaveSecurityGroup",
    "value" : null
  }, {
    "name" : "amazonEMRConfig.instanceCount",
    "value" : 2
  }, {
    "name" : "amazonEMRConfig.masterInstanceType",
    "value" : null
  }, {
    "name" : "amazonEMRConfig.masterInstanceTypeCustom",
    "value" : null
  }, {
    "name" : "amazonEMRConfig.slaveInstanceType",
    "value" : null
  }, {
    "name" : "amazonEMRConfig.slaveInstanceTypeCustom",
    "value" : null
  }, {
    "name" : "transformerEMRConfig.userRegion",
    "value" : null
  }, {
    "name" : "transformerEMRConfig.userRegionCustom",
    "value" : null
  }, {
    "name" : "transformerEMRConfig.useIAMRoles",
    "value" : false
  }, {
    "name" : "transformerEMRConfig.accessKey",
    "value" : ""
  }, {
    "name" : "transformerEMRConfig.secretKey",
    "value" : ""
  }, {
    "name" : "transformerEMRConfig.s3StagingUri",

```

```

    "value" : null
  }, {
    "name" : "transformerEMRConfig.provisionNewCluster",
    "value" : false
  }, {
    "name" : "transformerEMRConfig.clusterId",
    "value" : null
  }, {
    "name" : "transformerEMRConfig.emrVersion",
    "value" : null
  }, {
    "name" : "transformerEMRConfig.clusterPrefix",
    "value" : null
  }, {
    "name" : "transformerEMRConfig.terminateCluster",
    "value" : false
  }, {
    "name" : "transformerEMRConfig.loggingEnabled",
    "value" : true
  }, {
    "name" : "transformerEMRConfig.s3LogUri",
    "value" : null
  }, {
    "name" : "transformerEMRConfig.serviceRole",
    "value" : "EMR_DefaultRole"
  }, {
    "name" : "transformerEMRConfig.jobFlowRole",
    "value" : "EMR_EC2_DefaultRole"
  }, {
    "name" : "transformerEMRConfig.visibleToAllUsers",
    "value" : true
  }, {
    "name" : "transformerEMRConfig.ec2SubnetId",
    "value" : null
  }, {
    "name" : "transformerEMRConfig.masterSecurityGroup",
    "value" : null
  }, {
    "name" : "transformerEMRConfig.slaveSecurityGroup",
    "value" : null
  }, {
    "name" : "transformerEMRConfig.instanceCount",
    "value" : 2
  }, {
    "name" : "transformerEMRConfig.masterInstanceType",
    "value" : null
  }, {
    "name" : "transformerEMRConfig.masterInstanceTypeCustom",
    "value" : null
  }, {
    "name" : "transformerEMRConfig.slaveInstanceType",
    "value" : null
  }, {
    "name" : "transformerEMRConfig.slaveInstanceTypeCustom",
    "value" : null
  }

```

```

    } ],
    "uiInfo" : {
      "previewConfig" : {
        "previewSource" : "CONFIGURED_SOURCE",
        "batchSize" : 10,
        "timeout" : 30000,
        "writeToDestinations" : false,
        "executeLifecycleEvents" : false,
        "showHeader" : false,
        "showFieldType" : true,
        "rememberMe" : false
      }
    },
    "fragments" : [ ],
    "stages" : [ {
      "instanceName" : "Directory_01",
      "library" : "streamsets-datacollector-basic-lib",
      "stageName" : "com_streamsets_pipeline_stage_origin_spooldir_SpoolDirDSource",
      "stageVersion" : "11",
      "configuration" : [ {
        "name" : "conf.dataFormatConfig.compression",
        "value" : "NONE"
      }, {
        "name" : "conf.dataFormatConfig.filePatternInArchive",
        "value" : "*"
      }, {
        "name" : "conf.dataFormatConfig.charset",
        "value" : "UTF-8"
      }, {
        "name" : "conf.dataFormatConfig.removeCtrlChars",
        "value" : false
      }, {
        "name" : "conf.dataFormatConfig.textMaxLineLen",
        "value" : 1024
      }, {
        "name" : "conf.dataFormatConfig.useCustomDelimiter",
        "value" : false
      }, {
        "name" : "conf.dataFormatConfig.customDelimiter",
        "value" : "\\r\\n"
      }, {
        "name" : "conf.dataFormatConfig.includeCustomDelimiterInTheText",
        "value" : false
      }, {
        "name" : "conf.dataFormatConfig.jsonContent",
        "value" : "MULTIPLE_OBJECTS"
      }, {
        "name" : "conf.dataFormatConfig.jsonMaxObjectLen",
        "value" : 4096
      }, {
        "name" : "conf.dataFormatConfig.csvFileFormat",
        "value" : "CSV"
      }, {
        "name" : "conf.dataFormatConfig.csvHeader",

```

```

    "value" : "WITH_HEADER"
  }, {
    "name" : "conf.dataFormatConfig.csvAllowExtraColumns",
    "value" : false
  }, {
    "name" : "conf.dataFormatConfig.csvExtraColumnPrefix",
    "value" : "_extra_"
  }, {
    "name" : "conf.dataFormatConfig.csvMaxObjectLen",
    "value" : 1024
  }, {
    "name" : "conf.dataFormatConfig.csvCustomDelimiter",
    "value" : "|"
  }, {
    "name" : "conf.dataFormatConfig.multiCharacterFieldDelimiter",
    "value" : "||"
  }, {
    "name" : "conf.dataFormatConfig.multiCharacterLineDelimiter",
    "value" : "${str:unescapeJava('\\\\n')}"
  }, {
    "name" : "conf.dataFormatConfig.csvCustomEscape",
    "value" : "\\\"
  }, {
    "name" : "conf.dataFormatConfig.csvCustomQuote",
    "value" : "\""
  }, {
    "name" : "conf.dataFormatConfig.csvEnableComments",
    "value" : false
  }, {
    "name" : "conf.dataFormatConfig.csvCommentMarker",
    "value" : "#"
  }, {
    "name" : "conf.dataFormatConfig.csvIgnoreEmptyLines",
    "value" : true
  }, {
    "name" : "conf.dataFormatConfig.csvRecordType",
    "value" : "LIST_MAP"
  }, {
    "name" : "conf.dataFormatConfig.csvSkipStartLines",
    "value" : 0
  }, {
    "name" : "conf.dataFormatConfig.parseNull",
    "value" : false
  }, {
    "name" : "conf.dataFormatConfig.nullConstant",
    "value" : "\\\"\\\"\\N"
  }, {
    "name" : "conf.dataFormatConfig.xmlRecordElement",
    "value" : null
  }, {
    "name" : "conf.dataFormatConfig.preserveRootElement",
    "value" : true
  }, {
    "name" : "conf.dataFormatConfig.includeFieldXpathAttributes",
    "value" : false

```

```

}, {
  "name" : "conf.dataFormatConfig.xpathNamespaceContext",
  "value" : [ ]
}, {
  "name" : "conf.dataFormatConfig.outputFieldAttributes",
  "value" : false
}, {
  "name" : "conf.dataFormatConfig.xmlMaxObjectLen",
  "value" : 4096
}, {
  "name" : "conf.dataFormatConfig.logMode",
  "value" : "COMMON_LOG_FORMAT"
}, {
  "name" : "conf.dataFormatConfig.logMaxObjectLen",
  "value" : 1024
}, {
  "name" : "conf.dataFormatConfig.retainOriginalLine",
  "value" : false
}, {
  "name" : "conf.dataFormatConfig.customLogFormat",
  "value" : "%h %l %u %t \"%r\" %>s %b"
}, {
  "name" : "conf.dataFormatConfig.regex",
  "value" : "^(\\S+) (\\S+) (\\S+) \\[([\\w:/]+\\s[+\\-]\\d{4})\\] \\[\\S+\\] (\\S+) (\\S+) (\\S+)\\\" (\\d{3}) (\\d+)\"
", {
  "name" : "conf.dataFormatConfig.fieldPathsToGroupName",
  "value" : [ {
    "fieldPath" : "/",
    "group" : 1
  } ]
}, {
  "name" : "conf.dataFormatConfig.grokPatternDefinition",
  "value" : null
}, {
  "name" : "conf.dataFormatConfig.grokPattern",
  "value" : "%{COMMONAPACHELOG}"
}, {
  "name" : "conf.dataFormatConfig.onParseError",
  "value" : "ERROR"
}, {
  "name" : "conf.dataFormatConfig.maxStackTraceLines",
  "value" : 50
}, {
  "name" : "conf.dataFormatConfig.enableLog4jCustomLogFormat",
  "value" : false
}, {
  "name" : "conf.dataFormatConfig.log4jCustomLogFormat",
  "value" : "%r [%t] %-5p %c %x - %m%n"
}, {
  "name" : "conf.dataFormatConfig.avroSchema",
  "value" : null
}, {
  "name" : "conf.dataFormatConfig.schemaRegistryUrls",
  "value" : [ ]

```



```

}, {
  "name" : "conf.dataFormatConfig.basicAuth",
  "value" : ""
}, {
  "name" : "conf.dataFormatConfig.schemaLookupMode",
  "value" : "SUBJECT"
}, {
  "name" : "conf.dataFormatConfig.subject",
  "value" : null
}, {
  "name" : "conf.dataFormatConfig.schemaId",
  "value" : null
}, {
  "name" : "conf.dataFormatConfig.avroSkipUnionIndex",
  "value" : false
}, {
  "name" : "conf.dataFormatConfig.protoDescriptorFile",
  "value" : null
}, {
  "name" : "conf.dataFormatConfig.messageType",
  "value" : null
}, {
  "name" : "conf.dataFormatConfig.isDelimited",
  "value" : true
}, {
  "name" : "conf.dataFormatConfig.binaryMaxObjectLen",
  "value" : 1024
}, {
  "name" : "conf.dataFormatConfig.datagramMode",
  "value" : "SYSLOG"
}, {
  "name" : "conf.dataFormatConfig.typesDbPath",
  "value" : null
}, {
  "name" : "conf.dataFormatConfig.convertTime",
  "value" : false
}, {
  "name" : "conf.dataFormatConfig.excludeInterval",
  "value" : true
}, {
  "name" : "conf.dataFormatConfig.authFilePath",
  "value" : null
}, {
  "name" : "conf.dataFormatConfig.netflowOutputValuesMode",
  "value" : "RAW_AND_INTERPRETED"
}, {
  "name" : "conf.dataFormatConfig.maxTemplateCacheSize",
  "value" : -1
}, {
  "name" : "conf.dataFormatConfig.templateCacheTimeoutMs",
  "value" : -1
}, {
  "name" : "conf.dataFormatConfig.netflowOutputValuesModeDatagram",
  "value" : "RAW_AND_INTERPRETED"
}, {

```

```

    "name" : "conf.dataFormatConfig.maxTemplateCacheSizeDatagram",
    "value" : -1
  }, {
    "name" : "conf.dataFormatConfig.templateCacheTimeoutMsDatagram",
    "value" : -1
  }, {
    "name" : "conf.dataFormatConfig.wholeFileMaxObjectLen",
    "value" : 8192
  }, {
    "name" : "conf.dataFormatConfig.rateLimit",
    "value" : "-1"
  }, {
    "name" : "conf.dataFormatConfig.excelHeader",
    "value" : null
  }, {
    "name" : "conf.dataFormatConfig.excelSkipCellsWithNoHeader",
    "value" : false
  }, {
    "name" : "conf.dataFormatConfig.excelReadAllSheets",
    "value" : true
  }, {
    "name" : "conf.dataFormatConfig.excelSheetNames",
    "value" : [ ]
  }, {
    "name" : "conf.dataFormat",
    "value" : "DELIMITED"
  }, {
    "name" : "conf.spoolDir",
    "value" : "/opt/data/tutorial/origin"
  }, {
    "name" : "conf.numberOfThreads",
    "value" : 1
  }, {
    "name" : "conf.pathMatcherMode",
    "value" : "GLOB"
  }, {
    "name" : "conf.filePattern",
    "value" : "nyc_taxi*.csv"
  }, {
    "name" : "conf.useLastModified",
    "value" : "LEXICOGRAPHICAL"
  }, {
    "name" : "conf.processSubdirectories",
    "value" : false
  }, {
    "name" : "conf.allowLateDirectory",
    "value" : false
  }, {
    "name" : "conf.overrunLimit",
    "value" : 128
  }, {
    "name" : "conf.batchSize",
    "value" : 1000
  }, {
    "name" : "conf.poolingTimeoutSecs",

```

```

        "value" : 60
    }, {
        "name" : "conf.maxSpoolFiles",
        "value" : 1000
    }, {
        "name" : "conf.spoolingPeriod",
        "value" : 5
    }, {
        "name" : "conf.initialFileToProcess",
        "value" : null
    }, {
        "name" : "conf.errorArchiveDir",
        "value" : null
    }, {
        "name" : "conf.postProcessing",
        "value" : "NONE"
    }, {
        "name" : "conf.archiveDir",
        "value" : null
    }, {
        "name" : "conf.retentionTimeMins",
        "value" : 0
    }, {
        "name" : "stageOnRecordError",
        "value" : "TO_ERROR"
    } ],
    "uiInfo" : {
        "yPos" : 51,
        "stageType" : "SOURCE",
        "rawSource" : {
            "configuration" : [ {
                "name" : "fileName"
            } ]
        },
        "icon" : "directory.png",
        "description" : "",
        "label" : "Source",
        "xPos" : 61
    },
    "inputLanes" : [ ],
    "outputLanes" : [ "Directory_01OutputLane15886612861990" ],
    "eventLanes" : [ ],
    "services" : [ ]
}, {
    "instanceName" : "HadoopFS_01",
    "library" : "streamsets-datacollector-cdh_6_2-lib",
    "stageName" :
"com_streamsets_pipeline_stage_destination_hdfs_HdfsDTarget",
    "stageVersion" : "5",
    "configuration" : [ {
        "name" : "hdfsTargetConfigBean.uniquePrefix",
        "value" : "sdc-${sdc.id()}"
    }, {
        "name" : "hdfsTargetConfigBean.fileNameSuffix",
        "value" : null
    } ]
}

```

```

}, {
  "name" : "hdfsTargetConfigBean.dirPathTemplateInHeader",
  "value" : false
}, {
  "name" : "hdfsTargetConfigBean.dirPathTemplate",
  "value" : "/tmp/out/${YYYY()}}-${MM()}-${DD()}-${hh()}"
}, {
  "name" : "hdfsTargetConfigBean.timeZoneID",
  "value" : "UTC"
}, {
  "name" : "hdfsTargetConfigBean.timeDriver",
  "value" : "${time:now()}"
}, {
  "name" : "hdfsTargetConfigBean.maxRecordsPerFile",
  "value" : 0
}, {
  "name" : "hdfsTargetConfigBean.maxFileSize",
  "value" : 0
}, {
  "name" : "hdfsTargetConfigBean.idleTimeout",
  "value" : "${1 * HOURS}"
}, {
  "name" : "hdfsTargetConfigBean.compression",
  "value" : "NONE"
}, {
  "name" : "hdfsTargetConfigBean.otherCompression",
  "value" : null
}, {
  "name" : "hdfsTargetConfigBean.fileType",
  "value" : "TEXT"
}, {
  "name" : "hdfsTargetConfigBean.keyEl",
  "value" : "${uuid()}"
}, {
  "name" : "hdfsTargetConfigBean.seqFileCompressionType",
  "value" : "BLOCK"
}, {
  "name" : "hdfsTargetConfigBean.lateRecordsLimit",
  "value" : "${1 * HOURS}"
}, {
  "name" : "hdfsTargetConfigBean.rollIfHeader",
  "value" : false
}, {
  "name" : "hdfsTargetConfigBean.rollHeaderName",
  "value" : "roll"
}, {
  "name" : "hdfsTargetConfigBean.lateRecordsAction",
  "value" : "SEND_TO_ERROR"
}, {
  "name" : "hdfsTargetConfigBean.lateRecordsDirPathTemplate",
  "value" : "/tmp/late/${YYYY()}-${MM()}-${DD()}"
}, {
  "name" : "hdfsTargetConfigBean.dataFormat",
  "value" : "JSON"
}, {

```

```

        "name" : "hdfsTargetConfigBean.hdfsPermissionCheck",
        "value" : true
    }, {
        "name" : "hdfsTargetConfigBean.permissionEL",
        "value" : null
    }, {
        "name" : "hdfsTargetConfigBean.skipOldTempFileRecovery",
        "value" : false
    }, {
        "name" : "hdfsTargetConfigBean.dataGeneratorFormatConfig.charset",
        "value" : "UTF-8"
    }, {
        "name"
        :
        "hdfsTargetConfigBean.dataGeneratorFormatConfig.csvFileFormat",
        "value" : "CSV"
    }, {
        "name" : "hdfsTargetConfigBean.dataGeneratorFormatConfig.csvHeader",
        "value" : "NO_HEADER"
    }, {
        "name"
        :
        "hdfsTargetConfigBean.dataGeneratorFormatConfig.csvReplaceNewLines",
        "value" : true
    }, {
        "name"
        :
        "hdfsTargetConfigBean.dataGeneratorFormatConfig.csvReplaceNewLinesString",
        "value" : " "
    }, {
        "name"
        :
        "hdfsTargetConfigBean.dataGeneratorFormatConfig.csvCustomDelimiter",
        "value" : "|"
    }, {
        "name"
        :
        "hdfsTargetConfigBean.dataGeneratorFormatConfig.csvCustomEscape",
        "value" : "\\\"
    }, {
        "name"
        :
        "hdfsTargetConfigBean.dataGeneratorFormatConfig.csvCustomQuote",
        "value" : "\""
    }, {
        "name"
        :
        "hdfsTargetConfigBean.dataGeneratorFormatConfig.csvQuoteMode",
        "value" : "MINIMAL"
    }, {
        "name" : "hdfsTargetConfigBean.dataGeneratorFormatConfig.jsonMode",
        "value" : "MULTIPLE_OBJECTS"
    }, {
        "name"
        :
        "hdfsTargetConfigBean.dataGeneratorFormatConfig.textFieldPath",
        "value" : "/text"
    }, {
        "name"
        :
        "hdfsTargetConfigBean.dataGeneratorFormatConfig.textRecordSeparator",
        "value" : "\\n"
    }, {

```

```

        "name" : "hdfsTargetConfigBean.dataGeneratorFormatConfig.textFieldMissingAction",
        "value" : "ERROR"
    }, {
        "name" : "hdfsTargetConfigBean.dataGeneratorFormatConfig.textEmptyLineIfNull",
        "value" : false
    }, {
        "name" : "hdfsTargetConfigBean.dataGeneratorFormatConfig.avroSchemaSource",
        "value" : null
    }, {
        "name" : "hdfsTargetConfigBean.dataGeneratorFormatConfig.avroSchema",
        "value" : null
    }, {
        "name" : "hdfsTargetConfigBean.dataGeneratorFormatConfig.registerSchema",
        "value" : false
    }, {
        "name" : "hdfsTargetConfigBean.dataGeneratorFormatConfig.schemaRegistryUrlsForRegistration",
        "value" : [ ]
    }, {
        "name" : "hdfsTargetConfigBean.dataGeneratorFormatConfig.schemaRegistryUrls",
        "value" : [ ]
    }, {
        "name" : "hdfsTargetConfigBean.dataGeneratorFormatConfig.basicAuthUserInfo",
        "value" : ""
    }, {
        "name" : "hdfsTargetConfigBean.dataGeneratorFormatConfig.schemaLookupMode",
        "value" : "SUBJECT"
    }, {
        "name" : "hdfsTargetConfigBean.dataGeneratorFormatConfig.subject",
        "value" : null
    }, {
        "name" : "hdfsTargetConfigBean.dataGeneratorFormatConfig.subjectToRegister",
        "value" : null
    }, {
        "name" : "hdfsTargetConfigBean.dataGeneratorFormatConfig.basicAuthUserInfoForRegistration",
        "value" : ""
    }, {
        "name" : "hdfsTargetConfigBean.dataGeneratorFormatConfig.schemaId",
        "value" : null
    }, {
        "name" : "hdfsTargetConfigBean.dataGeneratorFormatConfig.avroCompression",
        "value" : "NULL"
    }, {

```

```

        "name" :
"hdFsTargetConfigBean.dataGeneratorFormatConfig.binaryFieldPath",
        "value" : "/"
    }, {
        "name" :
"hdFsTargetConfigBean.dataGeneratorFormatConfig.protoDescriptorFile",
        "value" : null
    }, {
        "name" : "hdFsTargetConfigBean.dataGeneratorFormatConfig.messageType",
        "value" : null
    }, {
        "name" : "hdFsTargetConfigBean.dataGeneratorFormatConfig.fileNameEL",
        "value" : null
    }, {
        "name" :
"hdFsTargetConfigBean.dataGeneratorFormatConfig.wholeFileExistsAction",
        "value" : "TO_ERROR"
    }, {
        "name" :
"hdFsTargetConfigBean.dataGeneratorFormatConfig.includeChecksumInTheEvents",
        "value" : false
    }, {
        "name" :
"hdFsTargetConfigBean.dataGeneratorFormatConfig.checksumAlgorithm",
        "value" : "MD5"
    }, {
        "name" :
"hdFsTargetConfigBean.dataGeneratorFormatConfig.xmlPrettyPrint",
        "value" : true
    }, {
        "name" :
"hdFsTargetConfigBean.dataGeneratorFormatConfig.xmlValidateSchema",
        "value" : false
    }, {
        "name" : "hdFsTargetConfigBean.dataGeneratorFormatConfig.xmlSchema",
        "value" : null
    }, {
        "name" : "hdFsTargetConfigBean.hdfsUri",
        "value" : ""
    }, {
        "name" : "hdFsTargetConfigBean.hdfsUser",
        "value" : ""
    }, {
        "name" : "hdFsTargetConfigBean.hdfsKerberos",
        "value" : false
    }, {
        "name" : "hdFsTargetConfigBean.hdfsConfDir",
        "value" : "/opt/streamsets-datacollector-3.15.0/resources/hadoop-conf"
    }, {
        "name" : "hdFsTargetConfigBean.hdfsConfigs",
        "value" : [ { } ]
    }, {
        "name" : "stageOnRecordError",
        "value" : "TO_ERROR"
    }, {

```

```

    "name" : "stageRequiredFields",
    "value" : [ ]
  }, {
    "name" : "stageRecordPreconditions",
    "value" : [ ]
  } ],
  "uiInfo" : {
    "yPos" : 50,
    "stageType" : "TARGET",
    "icon" : "hdfs.png",
    "description" : "",
    "label" : "Destination FS",
    "xPos" : 281
  },
  "inputLanes" : [ "Directory_01OutputLane15886612861990" ],
  "outputLanes" : [ ],
  "eventLanes" : [ ],
  "services" : [ ]
} ],
"errorStage" : {
  "instanceName" : "WritetoFile_ErrorStage",
  "library" : "streamsets-datacollector-basic-lib",
  "stageName" :
"com_streamsets_pipeline_stage_destination_recordstolocalfilesystem_ToErrorLoc
alFSDTarget",
  "stageVersion" : "1",
  "configuration" : [ {
    "name" : "directory",
    "value" : "/opt/data/tutorial/error"
  }, {
    "name" : "uniquePrefix",
    "value" : "sdc-${sdc:id()}"
  }, {
    "name" : "rotationIntervalSecs",
    "value" : "${1 * HOURS}"
  }, {
    "name" : "maxFileSizeMbs",
    "value" : 512
  } ],
  "uiInfo" : {
    "yPos" : 50,
    "stageType" : "TARGET",
    "icon" : "",
    "description" : "",
    "label" : "Error Records - Write to File",
    "xPos" : 60
  },
  "inputLanes" : [ ],
  "outputLanes" : [ ],
  "eventLanes" : [ ],
  "services" : [ ]
},
"info" : {
  "pipelineId" : "SampleTest6929ae91-0b30-4a8e-b111-1d64c778ce09",
  "title" : "SampleTest",

```



```

    "description" : "test pipeline",
    "created" : 1588661193429,
    "lastModified" : 1589870732059,
    "creator" : "admin",
    "lastModifier" : "admin",
    "lastRev" : "0",
    "uuid" : "29e035a7-2da3-4506-9980-1e7e66f72031",
    "valid" : true,
    "metadata" : {
      "labels" : [ ]
    },
    "name" : "SampleTest6929ae91-0b30-4a8e-b111-1d64c778ce09",
    "sdcVersion" : "3.15.0",
    "sdcId" : "9e30ea65-8e8d-11ea-b9ff-49cf138b18cc"
  },
  "metadata" : {
    "labels" : [ ]
  },
  "statsAggregatorStage" : {
    "instanceName" : "statsAggregatorStageInstance",
    "library" : "streamsets-datacollector-basic-lib",
    "stageName" :
"com_streamsets_pipeline_stage_destination_devnull_StatsDpmDirectlyDTarget",
    "stageVersion" : "1",
    "configuration" : [ ],
    "uiInfo" : {
      "stageType" : "TARGET",
      "label" : "Stats Aggregator -Write Directly to Control Hub - statistics
are not aggregated across Data Collectors"
    },
    "inputLanes" : [ ],
    "outputLanes" : [ ],
    "eventLanes" : [ ],
    "services" : [ ]
  },
  "startEventStages" : [ {
    "instanceName" : "Discard_StartEventStage",
    "library" : "streamsets-datacollector-basic-lib",
    "stageName" :
"com_streamsets_pipeline_stage_destination_devnull_ToErrorNullDTarget",
    "stageVersion" : "1",
    "configuration" : [ ],
    "uiInfo" : {
      "yPos" : 50,
      "stageType" : "TARGET",
      "icon" : "",
      "description" : "",
      "label" : "Start Event - Discard",
      "xPos" : 60
    },
    "inputLanes" : [ ],
    "outputLanes" : [ ],
    "eventLanes" : [ ],
    "services" : [ ]
  } ],

```

```

    "stopEventStages" : [ {
      "instanceName" : "Discard_StopEventStage",
      "library" : "streamsets-datacollector-basic-lib",
      "stageName" : "Discard_StopEventStage",
      "stageVersion" : "1",
      "configuration" : [ ],
      "uiInfo" : {
        "yPos" : 50,
        "stageType" : "TARGET",
        "icon" : "",
        "description" : "",
        "label" : "Stop Event - Discard",
        "xPos" : 60
      },
      "inputLanes" : [ ],
      "outputLanes" : [ ],
      "eventLanes" : [ ],
      "services" : [ ]
    } ],
    "testOriginStage" : {
      "instanceName" : "Test Origin - Dev Raw Data Source",
      "library" : "streamsets-datacollector-dev-lib",
      "stageName" : "Test Origin - Dev Raw Data Source",
      "stageVersion" : "3",
      "configuration" : [ {
        "name" : "rawData",
        "value" : "{\n  \"f1\" : \"abc\", \n  \"f2\" : \"xyz\", \n  \"f3\" : \"lmn\", \n}"
      }, {
        "name" : "stopAfterFirstBatch",
        "value" : false
      }, {
        "name" : "eventData",
        "value" : null
      }, {
        "name" : "stageOnRecordError",
        "value" : "TO_ERROR"
      } ],
      "uiInfo" : {
        "stageType" : "SOURCE",
        "label" : "Test Origin - Dev Raw Data Source"
      },
      "inputLanes" : [ ],
      "outputLanes" : [ ],
      "eventLanes" : [ ],
      "services" : [ {
        "service" : "com.streamsets.pipeline.api.service.dataformats.DataFormatParserService",
        "serviceVersion" : 2,

```

```
"configuration" : [ {
  "name" : "displayFormats",
  "value" : "DELIMITED,JSON,LOG,SDC_JSON,TEXT,XML"
}, {
  "name" : "dataFormat",
  "value" : "JSON"
}, {
  "name" : "dataFormatConfig.compression",
  "value" : "NONE"
}, {
  "name" : "dataFormatConfig.filePatternInArchive",
  "value" : "*"
}, {
  "name" : "dataFormatConfig.charset",
  "value" : "UTF-8"
}, {
  "name" : "dataFormatConfig.removeCtrlChars",
  "value" : false
}, {
  "name" : "dataFormatConfig.textMaxLineLen",
  "value" : 1024
}, {
  "name" : "dataFormatConfig.useCustomDelimiter",
  "value" : false
}, {
  "name" : "dataFormatConfig.customDelimiter",
  "value" : "\\r\\n"
}, {
  "name" : "dataFormatConfig.includeCustomDelimiterInTheText",
  "value" : false
}, {
  "name" : "dataFormatConfig.jsonContent",
  "value" : "MULTIPLE_OBJECTS"
}, {
  "name" : "dataFormatConfig.jsonMaxObjectLen",
  "value" : 4096
}, {
  "name" : "dataFormatConfig.csvFileFormat",
  "value" : "CSV"
}, {
  "name" : "dataFormatConfig.csvHeader",
  "value" : "NO_HEADER"
}, {
  "name" : "dataFormatConfig.csvAllowExtraColumns",
  "value" : false
}, {
  "name" : "dataFormatConfig.csvExtraColumnPrefix",
  "value" : "_extra_"
}, {
  "name" : "dataFormatConfig.csvMaxObjectLen",
  "value" : 1024
}, {
  "name" : "dataFormatConfig.csvCustomDelimiter",
  "value" : "|"
}, {
```

```

    "name" : "dataFormatConfig.multiCharacterFieldDelimiter",
    "value" : "||"
  }, {
    "name" : "dataFormatConfig.multiCharacterLineDelimiter",
    "value" : "${str:unescapeJava('\\\n')}"
  }, {
    "name" : "dataFormatConfig.csvCustomEscape",
    "value" : "\\"
  }, {
    "name" : "dataFormatConfig.csvCustomQuote",
    "value" : "\""
  }, {
    "name" : "dataFormatConfig.csvEnableComments",
    "value" : false
  }, {
    "name" : "dataFormatConfig.csvCommentMarker",
    "value" : "#"
  }, {
    "name" : "dataFormatConfig.csvIgnoreEmptyLines",
    "value" : false
  }, {
    "name" : "dataFormatConfig.csvRecordType",
    "value" : "LIST_MAP"
  }, {
    "name" : "dataFormatConfig.csvSkipStartLines",
    "value" : 0
  }, {
    "name" : "dataFormatConfig.parseNull",
    "value" : false
  }, {
    "name" : "dataFormatConfig.nullConstant",
    "value" : "\\N"
  }, {
    "name" : "dataFormatConfig.xmlRecordElement",
    "value" : null
  }, {
    "name" : "dataFormatConfig.preserveRootElement",
    "value" : false
  }, {
    "name" : "dataFormatConfig.includeFieldXpathAttributes",
    "value" : false
  }, {
    "name" : "dataFormatConfig.xPathNamespaceContext",
    "value" : [ ]
  }, {
    "name" : "dataFormatConfig.outputFieldAttributes",
    "value" : false
  }, {
    "name" : "dataFormatConfig.xmlMaxObjectLen",
    "value" : 4096
  }, {
    "name" : "dataFormatConfig.logMode",
    "value" : "COMMON_LOG_FORMAT"
  }, {
    "name" : "dataFormatConfig.logMaxObjectLen",

```

```

    "value" : 1024
  }, {
    "name" : "dataFormatConfig.retainOriginalLine",
    "value" : false
  }, {
    "name" : "dataFormatConfig.customLogFormat",
    "value" : "%h %l %u %t \"%r\" %>s %b"
  }, {
    "name" : "dataFormatConfig.regex",
    "value" : "^(\\S+) (\\S+) (\\S+) \\[([\\w:/]+\\s[+\\-]\\d{4})\\] \\[\\S+\\] (\\S+) (\\S+)\\\" (\\d{3}) (\\d+)"
  }, {
    "name" : "dataFormatConfig.fieldPathsToGroupName",
    "value" : [ {
      "fieldPath" : "/",
      "group" : 1
    } ]
  }, {
    "name" : "dataFormatConfig.grokPatternDefinition",
    "value" : null
  }, {
    "name" : "dataFormatConfig.grokPattern",
    "value" : "%{COMMONAPACHELOG}"
  }, {
    "name" : "dataFormatConfig.onParseError",
    "value" : "ERROR"
  }, {
    "name" : "dataFormatConfig.maxStackTraceLines",
    "value" : 50
  }, {
    "name" : "dataFormatConfig.enableLog4jCustomLogFormat",
    "value" : false
  }, {
    "name" : "dataFormatConfig.log4jCustomLogFormat",
    "value" : "%r [%t] %-5p %c %x - %m%n"
  }, {
    "name" : "dataFormatConfig.avroSchemaSource",
    "value" : null
  }, {
    "name" : "dataFormatConfig.avroSchema",
    "value" : null
  }, {
    "name" : "dataFormatConfig.schemaRegistryUrls",
    "value" : [ ]
  }, {
    "name" : "dataFormatConfig.basicAuth",
    "value" : ""
  }, {
    "name" : "dataFormatConfig.schemaLookupMode",
    "value" : "SUBJECT"
  }, {
    "name" : "dataFormatConfig.subject",
    "value" : null
  }, {
    "name" : "dataFormatConfig.schemaId",

```

```

    "value" : null
  }, {
    "name" : "dataFormatConfig.avroSkipUnionIndex",
    "value" : false
  }, {
    "name" : "dataFormatConfig.protoDescriptorFile",
    "value" : null
  }, {
    "name" : "dataFormatConfig.messageType",
    "value" : null
  }, {
    "name" : "dataFormatConfig.isDelimited",
    "value" : false
  }, {
    "name" : "dataFormatConfig.binaryMaxObjectLen",
    "value" : 1024
  }, {
    "name" : "dataFormatConfig.datagramMode",
    "value" : "SYSLOG"
  }, {
    "name" : "dataFormatConfig.typesDbPath",
    "value" : null
  }, {
    "name" : "dataFormatConfig.convertTime",
    "value" : false
  }, {
    "name" : "dataFormatConfig.excludeInterval",
    "value" : false
  }, {
    "name" : "dataFormatConfig.authFilePath",
    "value" : null
  }, {
    "name" : "dataFormatConfig.netflowOutputValuesMode",
    "value" : "RAW_AND_INTERPRETED"
  }, {
    "name" : "dataFormatConfig.maxTemplateCacheSize",
    "value" : -1
  }, {
    "name" : "dataFormatConfig.templateCacheTimeoutMs",
    "value" : -1
  }, {
    "name" : "dataFormatConfig.netflowOutputValuesModeDatagram",
    "value" : "RAW_AND_INTERPRETED"
  }, {
    "name" : "dataFormatConfig.maxTemplateCacheSizeDatagram",
    "value" : -1
  }, {
    "name" : "dataFormatConfig.templateCacheTimeoutMsDatagram",
    "value" : -1
  }, {
    "name" : "dataFormatConfig.wholeFileMaxObjectLen",
    "value" : 8192
  }, {
    "name" : "dataFormatConfig.rateLimit",
    "value" : "-1"
  }

```

```

    }, {
      "name" : "dataFormatConfig.verifyChecksum",
      "value" : false
    }, {
      "name" : "dataFormatConfig.excelHeader",
      "value" : null
    }, {
      "name" : "dataFormatConfig.excelSkipCellsWithNoHeader",
      "value" : false
    }, {
      "name" : "dataFormatConfig.excelReadAllSheets",
      "value" : false
    }, {
      "name" : "dataFormatConfig.excelSheetNames",
      "value" : [ ]
    } ]
  } ]
},
"valid" : true,
"issues" : {
  "pipelineIssues" : [ ],
  "stageIssues" : { },
  "issueCount" : 0
},
"previewable" : true
},
"pipelineRules" : {
  "schemaVersion" : 3,
  "version" : 2,
  "metricsRuleDefinitions" : [ {
    "id" : "badRecordsAlertID",
    "alertText" : "High incidence of Error Records",
    "metricId" : "pipeline.batchErrorRecords.counter",
    "metricType" : "COUNTER",
    "metricElement" : "COUNTER_COUNT",
    "condition" : "${value() > 100}",
    "sendEmail" : false,
    "enabled" : false,
    "timestamp" : 1588661193477,
    "valid" : true
  }, {
    "id" : "stageErrorAlertID",
    "alertText" : "High incidence of Stage Errors",
    "metricId" : "pipeline.batchErrorMessages.counter",
    "metricType" : "COUNTER",
    "metricElement" : "COUNTER_COUNT",
    "condition" : "${value() > 100}",
    "sendEmail" : false,
    "enabled" : false,
    "timestamp" : 1588661193477,
    "valid" : true
  }, {
    "id" : "idleGaugeID",
    "alertText" : "Pipeline is Idle",
    "metricId" : "RuntimeStatsGauge.gauge",

```

```
    "metricType" : "GAUGE",
    "metricElement" : "TIME_OF_LAST_RECEIVED_RECORD",
    "condition" : "${time:now() - value() > 120000}",
    "sendEmail" : false,
    "enabled" : false,
    "timestamp" : 1588661193477,
    "valid" : true
  }, {
    "id" : "batchTimeAlertID",
    "alertText" : "Batch taking more time to process",
    "metricId" : "RuntimeStatsGauge.gauge",
    "metricType" : "GAUGE",
    "metricElement" : "CURRENT_BATCH_AGE",
    "condition" : "${value() > 200}",
    "sendEmail" : false,
    "enabled" : false,
    "timestamp" : 1588661193477,
    "valid" : true
  } ],
  "dataRuleDefinitions" : [ ],
  "driftRuleDefinitions" : [ ],
  "uuid" : "7aed0bef-c75c-498c-9f30-68ac657f899a",
  "configuration" : [ {
    "name" : "emailIDs",
    "value" : [ ]
  }, {
    "name" : "webhookConfigs",
    "value" : [ ]
  } ],
  "ruleIssues" : [ ],
  "configIssues" : [ ]
},
"libraryDefinitions" : null
}
```