```
{"paragraphs":[{"config":{"colWidth":12,"graph":
{"mode":"table","height":300,"optionOpen":false,"keys":[],"values":
[], "groups": [], "scatter": {}}, "enabled": true, "editorMode": "ace/mode/
scala"},"settings":{"params":{},"forms":
{}},"jobName":"paragraph 1486138474599 844983353","id":"20170203-11143
4_676543515","dateCreated":"2017-02-03T11:14:34-0500","status":"FINISH
ED", "progressUpdateIntervalMs":500, "$$hashKey": "object:231", "text": "//
Imports\nimport org.apache.spark.sql.functions. \nimport
org.joda.time.format.DateTimeFormat
\n","dateUpdated":"2017-02-03T11:14:55-0500","dateFinished":"2017-02-0
3T11:15:10-0500"."dateStarted":"2017-02-03T11:14:55-0500"."result":
{"code":"SUCCESS","type":"TEXT","msg":"\nimport
org.apache.spark.sql.functions._\n\nimport
org.joda.time.format.DateTimeFormat\n"},"focus":true},{"config":
{"colWidth":12,"graph":{"mode":"table","height":
300, "optionOpen": false, "keys": [], "values": [], "groups": [], "scatter":
{}},"enabled":true},"settings":{"params":{},"forms":
{}},"jobName":"paragraph_1486138495329_511421381","id":"20170203-11145
5_2068888952","dateCreated":"2017-02-03T11:14:55-0500","status":"FINIS
HED","progressUpdateIntervalMs":500,"focus":true,"$$hashKey":"object:
308", "dateUpdated": "2017-02-03T11:18:54-0500", "dateFinished": "2017-02-
03T11:18:54-0500", "dateStarted": "2017-02-03T11:18:54-0500", "result":
{"code": "SUCCESS", "type": "TEXT", "msg": "bash: wget: command not found
\nbash: line 1: wget: command not found\nbash: line 2: wget: command
not found\nbash: line 3: wget: command not found\ndownloaded
\n"},"text":"%sh\n\nwget http://stat-computing.org/dataexpo/
2009/2007.csv.bz2 -0 /tmp/flights 2007.csv.bz2\nwget http://stat-
computing.org/dataexpo/2009/2008.csv.bz2 -0 /tmp/
flights_2008.csv.bz2\nwget ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/
daily/by year/2007.csv.qz -0 /tmp/weather 2007.csv.qz\nwqet ftp://
ftp.ncdc.noaa.gov/pub/data/ghcn/daily/by_year/2008.csv.gz -0 /tmp/
weather_2008.csv.gz\necho \"downloaded\"\n\n"},{"config":{"colWidth":
12, "graph": {"mode": "table", "height": 300, "optionOpen": false, "keys":
[],"values":[],"groups":[],"scatter":{}},"enabled":true},"settings":
{"params":{},"forms":
{}},"jobName":"paragraph_1486138734122_-1365061946","id":"20170203-111
854 217040968", "dateCreated": "2017-02-03T11:18:54-0500", "status": "ERRO
R", "progressUpdateIntervalMs":500, "focus":true, "$$hashKey": "object:
394","dateUpdated":"2017-02-03T11:19:39-0500","dateFinished":"2017-02-
03T11:19:39-0500", "dateStarted": "2017-02-03T11:19:39-0500", "result":
{"code":"ERROR","type":"TEXT","msg":"bash: line 1: hadoop: command not
found\nbash: line 2: hadoop: command not found\nbash: line 5: hadoop:
command not found\nbash: line 6: hadoop: command not found\nbash: line
7: hadoop: command not found\nbash: line 1: hadoop: command not found
\nbash: line 2: hadoop: command not found\nbash: line 5: hadoop:
command not found\nbash: line 6: hadoop: command not found\nbash: line
7: hadoop: command not found\nExitValue: 127"},"text":"%sh\n\n#remove
existing copies of dataset from HDFS\nhadoop fs -rm -r -f /tmp/
airflightsdelays\nhadoop fs -mkdir /tmp/airflightsdelays\n\n#put data
into HDFS\nhadoop fs -put /tmp/flights 200*.bz2 /tmp/airflightsdelays/
```

```
\nhadoop fs -put /tmp/weather 200*.gz /tmp/airflightsdelays/\nhadoop
fs -ls -h /tmp/airflightsdelays/"},{"config":{"colWidth":12,"graph":
{"mode":"table","height":300,"optionOpen":false,"keys":[],"values":
[], "groups": [], "scatter": {}}, "enabled": true}, "settings": {"params":
{}."forms":
{}},"jobName":"paragraph_1486138779852_-73567158","id":"20170203-11193
9 1479170863", "dateCreated": "2017-02-03T11:19:39-0500", "status": "ERROR
","progressUpdateIntervalMs":500,"focus":true,"$$hashKey":"object:
479","dateUpdated":"2017-02-03T11:20:01-0500","dateFinished":"2017-02-
03T11:20:02-0500", "dateStarted": "2017-02-03T11:20:01-0500", "result":
{"code":"ERROR","type":"TEXT","msg":"Must be used before
SparkInterpreter (%spark) initialized\nHint: put this paragraph before
any Spark code and restart Zeppelin/Interpreter"},"text":"%dep\n
\nz.reset()\nz.load(\"joda-time:joda-time:2.9.1\")\n\n"},{"config":
{"colWidth":12,"graph":{"mode":"table","height":
300, "optionOpen": false, "keys": [], "values": [], "groups": [], "scatter":
{}},"enabled":true},"settings":{"params":{},"forms":
{}},"jobName":"paragraph_1486138801027_-575264399","id":"20170203-1120
01_881235926","dateCreated":"2017-02-03T11:20:01-0500","status":"FINIS
HED","progressUpdateIntervalMs":500,"focus":true,"$$hashKey":"object:
557", "dateUpdated": "2017-02-03T11:20:58-0500", "dateFinished": "2017-02-
03T11:21:02-0500", "dateStarted": "2017-02-03T11:20:58-0500", "result":
{"code":"SUCCESS","type":"TEXT","msg":"\nimport org.apache.spark.rdd._
\n\nimport scala.collection.JavaConverters._\n\nimport
au.com.bytecode.opencsv.CSVReader\n\nimport java.io._\n\nimport
org.joda.time._\n\nimport org.joda.time.format._\n\nimport
org.joda.time.format.DateTimeFormat\n\nimport org.joda.time.DateTime\n
\nimport org.joda.time.Days\n"},"text":"%spark\n\nimport
org.apache.spark.rdd._\nimport scala.collection.JavaConverters._
\nimport au.com.bytecode.opencsv.CSVReader\n\nimport java.io. \nimport
org.joda.time._\nimport org.joda.time.format._\nimport
org.joda.time.format.DateTimeFormat\nimport org.joda.time.DateTime
\nimport org.joda.time.Days"},{"config":{"colWidth":12,"graph":
{"mode":"table","height":300,"optionOpen":false,"keys":[],"values":
[], "groups":[], "scatter":{}}, "enabled":true}, "settings":{"params":
{},"forms":
{}},"jobName":"paragraph 1486138858680 1716592853","id":"20170203-1120
58_500848636","dateCreated":"2017-02-03T11:20:58-0500","status":"FINIS
HED","progressUpdateIntervalMs":500,"focus":true,"$$hashKey":"object:
638", "dateUpdated": "2017-02-03T11:24:20-0500", "dateFinished": "2017-02-
03T11:24:22-0500", "dateStarted": "2017-02-03T11:24:20-0500", "result":
{"code":"SUCCESS","type":"TEXT","msg":"\nwarning: Class org.joda.convert.FromString not found — continuing with a stub.\n
\nwarning: Class org.joda.convert.ToString not found - continuing with
a stub.\n\nwarning: Class org.joda.convert.ToString not found -
continuing with a stub.\n\nwarning: Class org.joda.convert.FromString
not found - continuing with a stub.\n\nwarning: Class
org.joda.convert.ToString not found - continuing with a stub.\n
\nwarning: Class org.joda.convert.FromString not found - continuing
with a stub.\n\nwarning: Class org.joda.convert.ToString not found -
```

```
continuing with a stub.\n\ndefined class DelayRec\n"},"text":"case
class DelayRec(year: String,\n\t\t\t\t\tmonth: String,\n\t\t\t\t
\tdayOfMonth: String,\n\t\t\t\t\tdayOfWeek: String,\n\t\t\t\t
\tcrsDepTime: String,\n\t\t\t\t\tdepDelay: String,\n\t\t\t\torigin:
String,\n\t\t\tdistance: String,\n\t\t\tcancelled: String) {\n
\n \t val holidays = List(\"01/01/2007\", \"01/15/2007\", \)
\"02/19/2007\", \"05/28/2007\", \"06/07/2007\", \"07/04/2007\",\n\t\\"09/03/2007\", \"10/08/2007\",\"11/11/2007\", \"11/22/2007\",\"12/25/2007\",\n\t\\"01/01/2008\",\"01/21/2008\",\"02/18/2008\"
                      \"01/01/2008\", \"01/21/2008\", \"02/18/2008\",
\"05/22/2008\", \"05/26/2008\", \"07/04/2008\",\n\t
                                                        \"09/01/2008\"
\"10/13/2008\",\"11/11/2008\",\"11/27/2008\",\"12/25/2008\")\n\t\n
def gen_features: (String, Array[Double]) = {\n
                                                          val values =
                     depDelay.toDouble,\n
Array(\n
                                                       month.toDouble,\n
dayOfMonth.toDouble.
                                      dayOfWeek.toDouble,
get_hour(crsDepTime).toDouble, \n
                                                distance.toDouble, \n
days_from_nearest_holiday(year.toInt, month.toInt, dayOfMonth.toInt)
                        new Tuple2(to date(year.toInt, month.toInt,
\n
          )
dayOfMonth.toInt), values) \n
                                   }\n\n
                                             def get hour(depTime:
String) : String = \"%04d\".format(depTime.toInt).take(2)
                                                                    def
to_date(year: Int, month: Int, day: Int) = \"%04d%02d%02d
\".format(year, month, day) \n \n
                                         def
days_from_nearest_holiday(year:Int, month:Int, day:Int): Int = { \n
val sampleDate = new org.joda.time.DateTime(year, month, day, 0, 0)\n
          holidays.foldLeft(3000) { (r, c) \Rightarrow \n
                                                            val holiday =
org.joda.time.format.DateTimeFormat.forPattern(\"MM/dd/yyyy
\").parseDateTime(c) \n
                                 val distance =
Math.abs(org.joda.time.Days.daysBetween(holiday, sampleDate).getDays)
          math.min(r, distance) \n
                                          } \n
                                                      \n} "},{"config":
{"colWidth":12,"graph":{"mode":"table","height":
300, "optionOpen": false, "keys": [], "values": [], "groups": [], "scatter":
{}},"enabled":true},"settings":{"params":{},"forms":
{}},"jobName":"paragraph 1486139060268 888303448","id":"20170203-11242
0_2146232883","dateCreated":"2017-02-03T11:24:20-0500","status":"ERROR
","progressUpdateIntervalMs":500,"focus":true,"$$hashKey":"object:
717", "dateUpdated": "2017-02-03T11:25:57-0500", "dateFinished": "2017-02-
03T11:25:57-0500", "dateStarted": "2017-02-03T11:25:57-0500", "result":
{"code":"ERROR","type":"TEXT","msg":"Table or view not found:
data_2007tmp; line 2 pos 5\nset zeppelin.spark.sql.stacktrace = true
to see full stacktrace"},"text":"%sql\n\nselect dayofWeek, case when
depDelay > 15 then 'delayed' else 'ok' end , count(1)\nfrom
data_2007tmp \ngroup by dayofweek , case when depDelay > 15 then
'delayed' else 'ok' end\n\n"},{"config":{"colWidth":12,"graph":
{"mode":"table","height":300,"optionOpen":false,"keys":[],"values":
[],"groups":[],"scatter":{}},"enabled":true},"settings":{"params":
{},"forms":
{}},"jobName":"paragraph_1486139133471_116420533","id":"20170203-11253
3_57758758","dateCreated":"2017-02-03T11:25:33-0500","status":"ERROR",
"progressUpdateIntervalMs":500,"focus":true,"$$hashKey":"object:
796", "dateUpdated": "2017-02-03T11:26:19-0500", "dateFinished": "2017-02-
03T11:26:19-0500","dateStarted":"2017-02-03T11:26:19-0500","result":
```

```
{"code": "ERROR", "type": "TEXT", "msg": "Table or view not found:
data 2007tmp; line 2 pos 6\nset zeppelin.spark.sql.stacktrace = true
to see full stacktrace"},"text":"%sql\n\nselect cast( cast(crsDepTime
as int) / 100 as int) as hour, case when depDelay > 15 then 'delayed'
else 'ok' end as delay, count(1) as count\nfrom data 2007tmp \ngroup
by cast(crsDepTime as int) / 100 as int), case when depDelay >
15 then 'delayed' else 'ok' end"},{"config":{"colWidth":12,"graph":
{"mode":"table", "height": 300, "optionOpen": false, "keys": [], "values":
[], "groups": [], "scatter": {}}, "enabled": true}, "settings": {"params":
{},"forms":
{}},"jobName":"paragraph_1486139179540_1323221302","id":"20170203-1126
19_1235516553","dateCreated":"2017-02-03T11:26:19-0500","status":"READ
Y", "progressUpdateIntervalMs":500, "focus":true, "$$hashKey": "object:
885", "dateUpdated": "2017-02-03T11:26:29-0500", "text": ""}], "name": "flig
ht1","id":"2CAZQU327","angularObjects":{"2C925EGRA:shared_process":
[],"2C9AR5P2X:shared_process":[],"2CAACWD6V:shared_process":
[],"2CA5UHH9B:shared_process":[],"2C9Q8K6HQ:shared_process":
[],"2C73N4D6Z:shared_process":[],"2CADZKEBT:shared_process":
[],"2C78DJQCW:shared_process":[],"2C8MPXM6P:shared_process":
[],"2C9JYPQ6M:shared_process":[],"2C8KAD5CR:shared_process":
[],"2CAZP7M7G:shared_process":[],"2CB35NKZD:shared_process":
[],"2C8K2JXZD:shared_process":[],"2C9CU3JPT:shared_process":
[],"2CB258YDW:shared_process":[],"2C7PVVVR2:shared_process":[],"2C7VY9H2F:shared_process":[]},"config":
{"looknfeel":"default"},"info":{}}
```