In [1]:

```
small_size = 1  # choose 0 or 1.
```

In [2]:

```
import pyrasterframes
from pyrasterframes.rasterfunctions import *

from pyspark.sql.functions import lit
import pyspark.sql.functions as F
import pandas as pd
```

In [3]:

```
from pyrasterframes.utils import create_rf_spark_session
spark = create_rf_spark_session(**{
    'spark.app.name': 'fire scale small' if small_size else 'fire scale LARGE',
    'spark.driver.memory': '8G',
    'spark.driver.extraJavaOptions':
        '-Dgeotrellis.raster.gdal.options.CPL_DEBUG=ON ' +
        '-Dgeotrellis.raster.gdal.useExceptions=false ' +
        '-Dgeotrellis.raster.gdal.options.GDAL_HTTP_MAX_RETRY=10 ' +
        '-Dgeotrellis.raster.gdal.options.GDAL_HTTP_RETRY_DELAY=2 '
})
```

In [4]:

```
from pyrasterframes.utils import gdal_version
gdal_version()
```

Out[4]:

```
'GDAL 3.0.4, released 2020/01/28'
```

In [5]:

```
from pyrasterframes.utils import build_info
build_info()
```

Out[5]:

```
{'scalaVersion': '2.11.12', 'sbtVersion': '1.3.7', 'name': 'core', 'rfS
parkVersion': '2.4.4', 'rfGeoMesaVersion': '2.2.1', 'GDAL': 'GDAL 3.0.
4, released 2020/01/28', 'rfGeoTrellisVersion': '3.2.1-astraea', 'versi
on': '0.9.0-astraea.5c372561.a'}
```

The configured Spark resouces

In [6]:

```
import pandas as pd
catalog = pd.read_csv('small_catalog.csv' if small_size else 'large_catalog.csv')
```

In [7]:

```
catalog
```

Out[7]:

	id_before	datetime_before	eod_grid_id	eo
0	S2A_OPER_MSI_L2A_TL_SGS__20190531T161101_A0205...	2019-05-31 13:45:21.658163+00:00	MGRS-22LEP	
1	S2A_OPER_MSI_L2A_TL_SGS__20190531T161101_A0205...	2019-05-31 13:45:21.658163+00:00	MGRS-22LEP	
2	S2A_OPER_MSI_L2A_TL_SGS__20190531T161101_A0205...	2019-05-31 13:45:21.658163+00:00	MGRS-22LEP	
3	S2A_OPER_MSI_L2A_TL_SGS__20190531T161101_A0205...	2019-05-31 13:45:21.658163+00:00	MGRS-22LEP	
4	S2A_OPER_MSI_L2A_TL_SGS__20190531T161101_A0205...	2019-05-31 13:45:21.658163+00:00	MGRS-22LEP	
5	S2A_OPER_MSI_L2A_TL_SGS__20190531T161101_A0205...	2019-05-31 13:45:21.658163+00:00	MGRS-22LEP	
6	S2A_OPER_MSI_L2A_TL_SGS__20190531T161101_A0205...	2019-05-31 13:45:21.658163+00:00	MGRS-22LEP	
7	S2A_OPER_MSI_L2A_TL_SGS__20190531T161101_A0205...	2019-05-31 13:45:21.658163+00:00	MGRS-22LEP	
8	S2A_OPER_MSI_L2A_TL_SGS__20190531T161101_A0205...	2019-05-31 13:45:21.658163+00:00	MGRS-22LEP	
9	S2A_OPER_MSI_L2A_TL_SGS__20190531T161101_A0205...	2019-05-31 13:45:21.658163+00:00	MGRS-22LEP	
10	S2A_OPER_MSI_L2A_TL_SGS__20190531T161101_A0205...	2019-05-31 13:45:21.658163+00:00	MGRS-22LEP	
11	S2A_OPER_MSI_L2A_TL_SGS__20190531T161101_A0205...	2019-05-31 13:45:21.658163+00:00	MGRS-22LEP	
12	S2A_OPER_MSI_L2A_TL_SGS__20190531T161101_A0205...	2019-05-31 13:45:21.658163+00:00	MGRS-22LEP	
13	S2A_OPER_MSI_L2A_TL_SGS__20190531T161101_A0205...	2019-05-31 13:45:21.658163+00:00	MGRS-22LEP	
14	S2A_OPER_MSI_L2A_TL_SGS__20190531T161101_A0205...	2019-05-31 13:45:21.658163+00:00	MGRS-22LEP	

Read the raster data with before and after columns in it.

In [8]:

```
df = spark.read.raster(catalog,
                       ['B08_60m_before', 'B12_60m_before',
                        'B08_60m_after',  'B12_60m_after'])
```

In [9]:

```
df.columns
```

Out[9]:

```
['B08_60m_before_path',
 'B12_60m_before_path',
 'B08_60m_after_path',
 'B12_60m_after_path',
 'B08_60m_before',
 'B12_60m_before',
 'B08_60m_after',
 'B12_60m_after',
 'id_before',
 'datetime_before',
 'eod_grid_id',
 'eo_cloud_cover_before',
 'id_after',
 'datetime_after',
 'eo_cloud_cover_after']
```

Compute the normalized burn ratio.

In [10]:

```
df2 = df.withColumn('nbr_before',
                    rf_normalized_difference('B08_60m_before', 'B12_60m_before')) \
        .withColumn('nbr_after',
                    rf_normalized_difference('B08_60m_after',  'B12_60m_after')) \
         .drop('B08_60m_before', 'B12_60m_before',
                         'B08_60m_after',  'B12_60m_after')
```

In [11]:

```
df2.toPandas()
```

```
---------------------------------------------------------------------------
Py4JJavaError                             Traceback (most recent call last)
<ipython-input-11-d0e5e1145736> in <module>
----> 1 df2.toPandas()

/opt/conda/lib/python3.7/site-packages/pyspark/sql/dataframe.py in toPandas(self)
   2141
   2142         # Below is toPandas without Arrow optimization.
-> 2143         pdf = pd.DataFrame.from_records(self.collect(), columns=self.columns)
   2144
   2145         dtype = {}

/opt/conda/lib/python3.7/site-packages/pyspark/sql/dataframe.py in collect(self)
    532         """
    533         with SCCallSiteSync(self._sc) as css:
--> 534             sock_info = self._jdf.collectToPython()
    535         return list(_load_from_socket(sock_info, BatchedSerializer(PickleSerializer())))
    536

/opt/conda/lib/python3.7/site-packages/py4j/java_gateway.py in __call__(self, *args)
   1255         answer = self.gateway_client.send_command(command)
   1256         return_value = get_return_value(
-> 1257             answer, self.gateway_client, self.target_id, self.name)
   1258
   1259         for temp_arg in temp_args:

/opt/conda/lib/python3.7/site-packages/pyspark/sql/utils.py in deco(*a, **kw)
     61     def deco(*a, **kw):
     62         try:
---> 63             return f(*a, **kw)
     64         except py4j.protocol.Py4JJavaError as e:
     65             s = e.java_exception.toString()

/opt/conda/lib/python3.7/site-packages/py4j/protocol.py in get_return_value(answer, gateway_client, target_id, name)
    326                 raise Py4JJavaError(
    327                     "An error occurred while calling {0}{1}{2}.\n".
--> 328                     format(target_id, ".", name), value)
    329             else:
    330                 raise Py4JError(

Py4JJavaError: An error occurred while calling o132.collectToPython.
: org.apache.spark.SparkException: Job aborted due to stage failure: Task 64 in stage 1.0 failed 1 times, most recent failure: Lost task 64.0
 in stage 1.0 (TID 198, localhost, executor driver): geotrellis.raster.gdal.MalformedDataException: Unable to construct a RasterExtent from the Transformation given. GDAL Error Code: 3
```

```
        at geotrellis.raster.gdal.GDALDataset$.rasterExtent$extension1
(GDALDataset.scala:143)
        at geotrellis.raster.gdal.GDALRasterSource.gridExtent$lzycomput
e(GDALRasterSource.scala:93)
        at geotrellis.raster.gdal.GDALRasterSource.gridExtent(GDALRaste
rSource.scala:93)
        at geotrellis.raster.RasterMetadata$class.cols(RasterMetadata.s
cala:52)
        at geotrellis.raster.RasterSource.cols(RasterSource.scala:44)
        at geotrellis.raster.RasterSource.cols(RasterSource.scala:44)
        at geotrellis.raster.Grid.dimensions(Grid.scala:26)
        at geotrellis.raster.gdal.GDALRasterSource$$anonfun$readBounds
$1.apply(GDALRasterSource.scala:105)
        at geotrellis.raster.gdal.GDALRasterSource$$anonfun$readBounds
$1.apply(GDALRasterSource.scala:105)
        at scala.collection.Iterator$$anon$12.nextCur(Iterator.scala:43
5)
        at scala.collection.Iterator$$anon$12.hasNext(Iterator.scala:44
1)
        at scala.collection.Iterator$$anon$12.next(Iterator.scala:445)
        at scala.collection.Iterator$$anon$11.next(Iterator.scala:410)
        at org.locationtech.rasterframes.ref.RFRasterSource.read(RFRast
erSource.scala:61)
        at org.locationtech.rasterframes.ref.RasterRef.realizedTile$lzy
compute(RasterRef.scala:57)
        at org.locationtech.rasterframes.ref.RasterRef.realizedTile(Ras
terRef.scala:55)
        at org.locationtech.rasterframes.ref.RasterRef$RasterRefTile.de
legate(RasterRef.scala:72)
        at org.locationtech.rasterframes.tiles.FixedDelegatingTile.comb
ine(FixedDelegatingTile.scala:31)
        at geotrellis.raster.Tile.dualCombine(Tile.scala:93)
        at geotrellis.raster.mapalgebra.local.LocalTileBinaryOp$class.a
pply(LocalTileBinaryOp.scala:56)
        at geotrellis.raster.mapalgebra.local.Subtract$.apply(Subtract.
scala:29)
        at geotrellis.raster.mapalgebra.local.SubtractMethods$class.loc
alSubtract(Subtract.scala:57)
        at geotrellis.raster.mapalgebra.local.Implicits$withTileLocalMe
thods.localSubtract(Implicits.scala:25)
        at org.locationtech.rasterframes.expressions.localops.Normalize
dDifference.op(NormalizedDifference.scala:47)
        at org.locationtech.rasterframes.expressions.BinaryRasterOp$cla
ss.nullSafeEval(BinaryRasterOp.scala:66)
        at org.locationtech.rasterframes.expressions.localops.Normalize
dDifference.nullSafeEval(NormalizedDifference.scala:44)
        at org.apache.spark.sql.catalyst.expressions.BinaryExpression.e
val(Expression.scala:484)
        at org.apache.spark.sql.catalyst.expressions.GeneratedClass$Spe
cificUnsafeProjection.writeFields_6_1$(Unknown Source)
        at org.apache.spark.sql.catalyst.expressions.GeneratedClass$Spe
cificUnsafeProjection.apply(Unknown Source)
        at org.apache.spark.sql.catalyst.expressions.GeneratedClass$Spe
cificUnsafeProjection.apply(Unknown Source)
        at scala.collection.Iterator$$anon$11.next(Iterator.scala:410)
        at org.apache.spark.sql.execution.SparkPlan$$anonfun$2.apply(Sp
arkPlan.scala:256)
```

```
        at org.apache.spark.sql.execution.SparkPlan$$anonfun$2.apply(Sp
arkPlan.scala:247)
        at org.apache.spark.rdd.RDD$$anonfun$mapPartitionsInternal$1$$a
nonfun$apply$24.apply(RDD.scala:836)
        at org.apache.spark.rdd.RDD$$anonfun$mapPartitionsInternal$1$$a
nonfun$apply$24.apply(RDD.scala:836)
        at org.apache.spark.rdd.MapPartitionsRDD.compute(MapPartitionsR
DD.scala:52)
        at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:3
24)
        at org.apache.spark.rdd.RDD.iterator(RDD.scala:288)
        at org.apache.spark.scheduler.ResultTask.runTask(ResultTask.sca
la:90)
        at org.apache.spark.scheduler.Task.run(Task.scala:123)
        at org.apache.spark.executor.Executor$TaskRunner$$anonfun$10.ap
ply(Executor.scala:408)
        at org.apache.spark.util.Utils$.tryWithSafeFinally(Utils.scala:
1360)
        at org.apache.spark.executor.Executor$TaskRunner.run(Executor.s
cala:414)
        at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPool
Executor.java:1142)
        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoo
lExecutor.java:617)
        at java.lang.Thread.run(Thread.java:745)

Driver stacktrace:
        at org.apache.spark.scheduler.DAGScheduler.org$apache$spark$sch
eduler$DAGScheduler$$failJobAndIndependentStages(DAGScheduler.scala:188
9)
        at org.apache.spark.scheduler.DAGScheduler$$anonfun$abortStage
$1.apply(DAGScheduler.scala:1877)
        at org.apache.spark.scheduler.DAGScheduler$$anonfun$abortStage
$1.apply(DAGScheduler.scala:1876)
        at scala.collection.mutable.ResizableArray$class.foreach(Resiza
bleArray.scala:59)
        at scala.collection.mutable.ArrayBuffer.foreach(ArrayBuffer.sca
la:48)
        at org.apache.spark.scheduler.DAGScheduler.abortStage(DAGSchedu
ler.scala:1876)
        at org.apache.spark.scheduler.DAGScheduler$$anonfun$handleTaskS
etFailed$1.apply(DAGScheduler.scala:926)
        at org.apache.spark.scheduler.DAGScheduler$$anonfun$handleTaskS
etFailed$1.apply(DAGScheduler.scala:926)
        at scala.Option.foreach(Option.scala:257)
        at org.apache.spark.scheduler.DAGScheduler.handleTaskSetFailed
(DAGScheduler.scala:926)
        at org.apache.spark.scheduler.DAGSchedulerEventProcessLoop.doOn
Receive(DAGScheduler.scala:2110)
        at org.apache.spark.scheduler.DAGSchedulerEventProcessLoop.onRe
ceive(DAGScheduler.scala:2059)
        at org.apache.spark.scheduler.DAGSchedulerEventProcessLoop.onRe
ceive(DAGScheduler.scala:2048)
        at org.apache.spark.util.EventLoop$$anon$1.run(EventLoop.scala:
49)
        at org.apache.spark.scheduler.DAGScheduler.runJob(DAGScheduler.
scala:737)
```

```
        at org.apache.spark.SparkContext.runJob(SparkContext.scala:206
1)
        at org.apache.spark.SparkContext.runJob(SparkContext.scala:208
2)
        at org.apache.spark.SparkContext.runJob(SparkContext.scala:210
1)
        at org.apache.spark.SparkContext.runJob(SparkContext.scala:212
6)
        at org.apache.spark.rdd.RDD$$anonfun$collect$1.apply(RDD.scala:
945)
        at org.apache.spark.rdd.RDDOperationScope$.withScope(RDDOperati
onScope.scala:151)
        at org.apache.spark.rdd.RDDOperationScope$.withScope(RDDOperati
onScope.scala:112)
        at org.apache.spark.rdd.RDD.withScope(RDD.scala:363)
        at org.apache.spark.rdd.RDD.collect(RDD.scala:944)
        at org.apache.spark.sql.execution.SparkPlan.executeCollect(Spar
kPlan.scala:299)
        at org.apache.spark.sql.Dataset$$anonfun$collectToPython$1.appl
y(Dataset.scala:3263)
        at org.apache.spark.sql.Dataset$$anonfun$collectToPython$1.appl
y(Dataset.scala:3260)
        at org.apache.spark.sql.Dataset$$anonfun$52.apply(Dataset.scal
a:3370)
        at org.apache.spark.sql.execution.SQLExecution$$anonfun$withNew
ExecutionId$1.apply(SQLExecution.scala:78)
        at org.apache.spark.sql.execution.SQLExecution$.withSQLConfProp
agated(SQLExecution.scala:125)
        at org.apache.spark.sql.execution.SQLExecution$.withNewExecutio
nId(SQLExecution.scala:73)
        at org.apache.spark.sql.Dataset.withAction(Dataset.scala:3369)
        at org.apache.spark.sql.Dataset.collectToPython(Dataset.scala:3
260)
        at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
        at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAcce
ssorImpl.java:62)
        at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMe
thodAccessorImpl.java:43)
        at java.lang.reflect.Method.invoke(Method.java:498)
        at py4j.reflection.MethodInvoker.invoke(MethodInvoker.java:244)
        at py4j.reflection.ReflectionEngine.invoke(ReflectionEngine.jav
a:357)
        at py4j.Gateway.invoke(Gateway.java:282)
        at py4j.commands.AbstractCommand.invokeMethod(AbstractCommand.j
ava:132)
        at py4j.commands.CallCommand.execute(CallCommand.java:79)
        at py4j.GatewayConnection.run(GatewayConnection.java:238)
        at java.lang.Thread.run(Thread.java:745)
Caused by: geotrellis.raster.gdal.MalformedDataException: Unable to con
struct a RasterExtent from the Transformation given. GDAL Error Code: 3
        at geotrellis.raster.gdal.GDALDataset$.rasterExtent$extension1
(GDALDataset.scala:143)
        at geotrellis.raster.gdal.GDALRasterSource.gridExtent$lzycomput
e(GDALRasterSource.scala:93)
        at geotrellis.raster.gdal.GDALRasterSource.gridExtent(GDALRaste
rSource.scala:93)
        at geotrellis.raster.RasterMetadata$class.cols(RasterMetadata.s
```

```
cala:52)
        at geotrellis.raster.RasterSource.cols(RasterSource.scala:44)
        at geotrellis.raster.RasterSource.cols(RasterSource.scala:44)
        at geotrellis.raster.Grid.dimensions(Grid.scala:26)
        at geotrellis.raster.gdal.GDALRasterSource$$anonfun$readBounds
$1.apply(GDALRasterSource.scala:105)
        at geotrellis.raster.gdal.GDALRasterSource$$anonfun$readBounds
$1.apply(GDALRasterSource.scala:105)
        at scala.collection.Iterator$$anon$12.nextCur(Iterator.scala:43
5)
        at scala.collection.Iterator$$anon$12.hasNext(Iterator.scala:44
1)
        at scala.collection.Iterator$$anon$12.next(Iterator.scala:445)
        at scala.collection.Iterator$$anon$11.next(Iterator.scala:410)
        at org.locationtech.rasterframes.ref.RFRasterSource.read(RFRast
erSource.scala:61)
        at org.locationtech.rasterframes.ref.RasterRef.realizedTile$lzy
compute(RasterRef.scala:57)
        at org.locationtech.rasterframes.ref.RasterRef.realizedTile(Ras
terRef.scala:55)
        at org.locationtech.rasterframes.ref.RasterRef$RasterRefTile.de
legate(RasterRef.scala:72)
        at org.locationtech.rasterframes.tiles.FixedDelegatingTile.comb
ine(FixedDelegatingTile.scala:31)
        at geotrellis.raster.Tile.dualCombine(Tile.scala:93)
        at geotrellis.raster.mapalgebra.local.LocalTileBinaryOp$class.a
pply(LocalTileBinaryOp.scala:56)
        at geotrellis.raster.mapalgebra.local.Subtract$.apply(Subtract.
scala:29)
        at geotrellis.raster.mapalgebra.local.SubtractMethods$class.loc
alSubtract(Subtract.scala:57)
        at geotrellis.raster.mapalgebra.local.Implicits$withTileLocalMe
thods.localSubtract(Implicits.scala:25)
        at org.locationtech.rasterframes.expressions.localops.Normalize
dDifference.op(NormalizedDifference.scala:47)
        at org.locationtech.rasterframes.expressions.BinaryRasterOp$cla
ss.nullSafeEval(BinaryRasterOp.scala:66)
        at org.locationtech.rasterframes.expressions.localops.Normalize
dDifference.nullSafeEval(NormalizedDifference.scala:44)
        at org.apache.spark.sql.catalyst.expressions.BinaryExpression.e
val(Expression.scala:484)
        at org.apache.spark.sql.catalyst.expressions.GeneratedClass$Spe
cificUnsafeProjection.writeFields_6_1$(Unknown Source)
        at org.apache.spark.sql.catalyst.expressions.GeneratedClass$Spe
cificUnsafeProjection.apply(Unknown Source)
        at org.apache.spark.sql.catalyst.expressions.GeneratedClass$Spe
cificUnsafeProjection.apply(Unknown Source)
        at scala.collection.Iterator$$anon$11.next(Iterator.scala:410)
        at org.apache.spark.sql.execution.SparkPlan$$anonfun$2.apply(Sp
arkPlan.scala:256)
        at org.apache.spark.sql.execution.SparkPlan$$anonfun$2.apply(Sp
arkPlan.scala:247)
        at org.apache.spark.rdd.RDD$$anonfun$mapPartitionsInternal$1$$a
nonfun$apply$24.apply(RDD.scala:836)
        at org.apache.spark.rdd.RDD$$anonfun$mapPartitionsInternal$1$$a
nonfun$apply$24.apply(RDD.scala:836)
        at org.apache.spark.rdd.MapPartitionsRDD.compute(MapPartitionsR
```

```
DD.scala:52)
        at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:3
24)
        at org.apache.spark.rdd.RDD.iterator(RDD.scala:288)
        at org.apache.spark.scheduler.ResultTask.runTask(ResultTask.sca
la:90)
        at org.apache.spark.scheduler.Task.run(Task.scala:123)
        at org.apache.spark.executor.Executor$TaskRunner$$anonfun$10.ap
ply(Executor.scala:408)
        at org.apache.spark.util.Utils$.tryWithSafeFinally(Utils.scala:
1360)
        at org.apache.spark.executor.Executor$TaskRunner.run(Executor.s
cala:414)
        at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPool
Executor.java:1142)
        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoo
lExecutor.java:617)
        ... 1 more
```