Markdown Export Test

```
def interleave2D(x):
    x &= 0xffffffff
    x = (x | (x << 16)) & 0xffff0000ffff
    x = (x | (x << 8)) & 0xff00ff00ff00ff
    x = (x | (x << 4)) & 0xf0f0f0f0f0f0f0f
    x = (x | (x << 2)) & 0x3333333333333333
    x = (x | (x << 1)) & 0x5555555555555555
    return x

def zorder2D(srcip, dstip):
    x = interleave2D(srcip)
    y = interleave2D(dstip)
    z = y | (x << 1)
    return z
```

Like the wikipedia example
Z-order curve

```
z = zorder2D(2, 3)
z
```

```
13
```

```
max_32 = (2**32) - 1

z = zorder2D(max_32, max_32)
z
```

```
18446744073709551615
```

```
max_64 = (2**64) - 1
max_64
```

```
18446744073709551615
```

```

```

```

```

```
from pyspark import SparkContext, SparkConf
from pyspark.sql import SparkSession
from pyspark.sql.types import *
from pyspark.sql.functions import *
from pyspark.sql.functions import udf
from decimal import *
import pyarrow.parquet as pq
from os import listdir
from os.path import isfile, join

# Register a Spark python UDF
```

```
zorder2DUdf = udf(zorder2D, LongType())

# Generate a dataset of random data
def gen(numRows = 10, numFiles = 1):
    df = spark.range(start=0, end=numRows, numPartitions=numFiles)
    df = ( df.select(
            df.id,
            (rand() * 10000).cast(LongType()).alias('src_ip'),
            (rand() * 10000).cast(LongType()).alias('dst_ip'),
            (rand() * 500).cast(ShortType()).alias('src_port'),
            (rand() * 500).cast(ShortType()).alias('dst_port')
        ))
    df = df.withColumn('zorder', zorder2DUdf(df.src_ip, df.dst_ip) )
    return df

#spark.stop()
# if spark :
#     spark.stop()

NUM_FILES = 16
spark = SparkSession.builder.master('local').config('spark.sql.shuffle.partitions', NUM_FILES).getOrCreate()
df = gen(numRows=10000, numFiles = NUM_FILES)
#df = df.sort('src_ip')
#df = df.sort('zorder')
df = df.repartitionByRange(NUM_FILES, 'src_ip')
#df = df.repartitionByRange(NUM_FILES, 'zorder')

df.show(truncate=False)
df.write.mode('overwrite').parquet('file:///tmp/test')
spark.stop()
```

```
+---+------+------+--------+--------+--------+
|id |src_ip|dst_ip|src_port|dst_port|zorder  |
+---+------+------+--------+--------+--------+
|28 |117   |3082  |18      |108     |5253734 |
|29 |207   |6554  |369     |352     |21094894|
|45 |390   |5102  |451     |393     |17290364|
|61 |586   |9067  |59      |457     |67974349|
|65 |513   |497   |4       |331     |611587  |
|67 |234   |2541  |127     |367     |4324569 |
|75 |175   |9312  |78      |120     |68197546|
|110|481   |3402  |133     |488     |5486662 |
|116|338   |5979  |259     |135     |18297677|
|135|315   |9999  |88      |92      |68618975|
|142|408   |2372  |307     |71      |4428432 |
|150|11    |5905  |196     |23      |18153867|
|164|592   |54    |265     |407     |534292  |
|185|317   |12    |283     |492     |133874  |
|274|206   |4336  |341     |488     |16840104|
|282|321   |3237  |484     |20      |5399571 |
|299|373   |656   |294     |416     |420642  |
|310|286   |8508  |293     |393     |67307512|
|312|355   |628   |28      |165     |408858  |
|316|420   |6136  |396     |274     |18341216|
+---+------+------+--------+--------+--------+
only showing top 20 rows
```

```python
import pandas as pd

row_group_counter = 0

stat_schema = ['row_group', 'column', 'min', 'max']
pdf = pd.DataFrame(columns=stat_schema)

parquetfiles = [f for f in listdir('/tmp/test/') if join('/tmp/test/', f).endswith('.parquet')]
parquetfiles.sort()

for f in parquetfiles:
    path = join('/tmp/test/', f)
    if path.endswith('.parquet'):
        metadata = pq.read_metadata(path)
        for r in range(metadata.num_row_groups):
            row_group = metadata.row_group(r)
            for c in range(metadata.num_columns):
                column = row_group.column(c)
                row = [row_group_counter, column.path_in_schema, column.statistics.min, column.statistics.max]
                pdf = pdf.append(pd.DataFrame([row], columns=stat_schema))
        row_group_counter += 1

pdf = pdf[(pdf.column == 'zorder') | (pdf.column == 'src_ip') | (pdf.column == 'dst_ip')].sort_values(by=['row_g
pdf
```

<div>

<style scoped>

.dataframe tbody tr th:only-of-type {

vertical-align: middle;

}

```css
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

</style>

<table border="1" class="dataframe">

<thead>

<tr style="text-align: right;">

<th></th>

<th>row_group</th>

<th>column</th>

<th>min</th>

<th>max</th>

</tr>

</thead>

<tbody>

<tr>

<th>0</th>

<td>0</td>

<td>dst_ip</td>

```
<td>4</td>
<td>9999</td>
</tr>
<tr>
<th>0</th>
<td>0</td>
<td>src_ip</td>
<td>1</td>
<td>602</td>
</tr>
<tr>
<th>0</th>
<td>0</td>
<td>zorder</td>
<td>5652</td>
<td>68973896</td>
</tr>
<tr>
<th>0</th>
<td>1</td>
<td>dst_ip</td>
<td>0</td>
<td>9996</td>
</tr>
<tr>
<th>0</th>
<td>1</td>
<td>src_ip</td>
<td>603</td>
<td>1206</td>
</tr>
<tr>
<th>0</th>
<td>1</td>
<td>zorder</td>
<td>538738</td>
<td>70569312</td>
</tr>
<tr>
<th>0</th>
<td>2</td>
<td>dst_ip</td>
<td>28</td>
<td>9999</td>
</tr>
<tr>
<th>0</th>
<td>2</td>
<td>src_ip</td>
<td>1208</td>
<td>1780</td>
```

```
</tr>
<tr>
<th>0</th>
<td>2</td>
<td>zorder</td>
<td>2155035</td>
<td>71099763</td>
</tr>
<tr>
<th>0</th>
<td>3</td>
<td>dst_ip</td>
<td>13</td>
<td>9989</td>
</tr>
<tr>
<th>0</th>
<td>3</td>
<td>src_ip</td>
<td>1781</td>
<td>2410</td>
</tr>
<tr>
<th>0</th>
<td>3</td>
<td>zorder</td>
<td>2669497</td>
<td>76969583</td>
</tr>
<tr>
<th>0</th>
<td>4</td>
<td>dst_ip</td>
<td>8</td>
<td>9996</td>
</tr>
<tr>
<th>0</th>
<td>4</td>
<td>src_ip</td>
<td>2411</td>
<td>3072</td>
</tr>
<tr>
<th>0</th>
<td>4</td>
<td>zorder</td>
<td>8536006</td>
<td>77562584</td>
</tr>
<tr>
```

```
<th>0</th>
<td>5</td>
<td>dst_ip</td>
<td>28</td>
<td>9983</td>
</tr>
<tr>
<th>0</th>
<td>5</td>
<td>src_ip</td>
<td>3073</td>
<td>3628</td>
</tr>
<tr>
<th>0</th>
<td>5</td>
<td>zorder</td>
<td>10487627</td>
<td>79432822</td>
</tr>
<tr>
<th>0</th>
<td>6</td>
<td>dst_ip</td>
<td>8</td>
<td>9975</td>
</tr>
<tr>
<th>0</th>
<td>6</td>
<td>src_ip</td>
<td>3630</td>
<td>4273</td>
</tr>
<tr>
<th>0</th>
<td>6</td>
<td>zorder</td>
<td>11024850</td>
<td>102023968</td>
</tr>
<tr>
<th>0</th>
<td>7</td>
<td>dst_ip</td>
<td>11</td>
<td>9999</td>
</tr>
<tr>
<th>0</th>
<td>7</td>
```

```
<td>src_ip</td>
<td>4274</td>
<td>4916</td>
</tr>
<tr>
<th>0</th>
<td>7</td>
<td>zorder</td>
<td>33595471</td>
<td>102598871</td>
</tr>
<tr>
<th>0</th>
<td>8</td>
<td>dst_ip</td>
<td>9</td>
<td>9996</td>
</tr>
<tr>
<th>0</th>
<td>8</td>
<td>src_ip</td>
<td>4922</td>
<td>5491</td>
</tr>
<tr>
<th>0</th>
<td>8</td>
<td>zorder</td>
<td>34222022</td>
<td>104228612</td>
</tr>
<tr>
<th>0</th>
<td>9</td>
<td>dst_ip</td>
<td>13</td>
<td>9993</td>
</tr>
<tr>
<th>0</th>
<td>9</td>
<td>src_ip</td>
<td>5493</td>
<td>6182</td>
</tr>
<tr>
<th>0</th>
<td>9</td>
<td>zorder</td>
<td>35814281</td>
```

<td>110382441</td>
</tr>
<tr>
<th>0</th>
<td>10</td>
<td>dst_ip</td>
<td>6</td>
<td>9912</td>
</tr>
<tr>
<th>0</th>
<td>10</td>
<td>src_ip</td>
<td>6184</td>
<td>6799</td>
</tr>
<tr>
<th>0</th>
<td>10</td>
<td>zorder</td>
<td>41957186</td>
<td>110904934</td>
</tr>
<tr>
<th>0</th>
<td>11</td>
<td>dst_ip</td>
<td>2</td>
<td>9991</td>
</tr>
<tr>
<th>0</th>
<td>11</td>
<td>src_ip</td>
<td>6801</td>
<td>7444</td>
</tr>
<tr>
<th>0</th>
<td>11</td>
<td>zorder</td>
<td>42523774</td>
<td>112514854</td>
</tr>
<tr>
<th>0</th>
<td>12</td>
<td>dst_ip</td>
<td>17</td>
<td>9971</td>
</tr>

```
<tr>
<th>0</th>
<td>12</td>
<td>src_ip</td>
<td>7445</td>
<td>8057</td>
</tr>
<tr>
<th>0</th>
<td>12</td>
<td>zorder</td>
<td>44174942</td>
<td>113141538</td>
</tr>
<tr>
<th>0</th>
<td>13</td>
<td>dst_ip</td>
<td>15</td>
<td>9995</td>
</tr>
<tr>
<th>0</th>
<td>13</td>
<td>src_ip</td>
<td>8059</td>
<td>8708</td>
</tr>
<tr>
<th>0</th>
<td>13</td>
<td>zorder</td>
<td>44727213</td>
<td>202834157</td>
</tr>
<tr>
<th>0</th>
<td>14</td>
<td>dst_ip</td>
<td>30</td>
<td>9991</td>
</tr>
<tr>
<th>0</th>
<td>14</td>
<td>src_ip</td>
<td>8709</td>
<td>9306</td>
</tr>
<tr>
<th>0</th>
```

```
<td>14</td>
<td>zorder</td>
<td>134749121</td>
<td>204562211</td>
</tr>
<tr>
<th>0</th>
<td>15</td>
<td>dst_ip</td>
<td>10</td>
<td>9996</td>
</tr>
<tr>
<th>0</th>
<td>15</td>
<td>src_ip</td>
<td>9307</td>
<td>9997</td>
</tr>
<tr>
<th>0</th>
<td>15</td>
<td>zorder</td>
<td>136342192</td>
<td>205411724</td>
</tr>
</tbody>
</table>
</div>
```

```
```

```
```