qparquet - casting.md 4/4/2019

## Outline

Previous examples used a very basic table. In reality not all datatypes will automatically map from python to

This notebook shows one example of casting data as it passing between the languages.

```
//Imports and setup
p)import warnings
p)warnings.filterwarnings("ignore")
p)import pandas as pd
p)import numpy as np
p)import pyarrow as pa
p)import pyarrow.parquet as pq
```

Create a sample table. This time one column is a datetime64

```
p)times=[np.datetime64('2012-06-30T21:00:00.000000000-0400')] * 4
p)table=pd.DataFrame(columns=['easy','time'])
p)table['time'] = times
p)table['easy'] = [1,2,3,4]
p)print(table)
```

```
easy time
0 1 2012-07-01 01:00:00
1 2 2012-07-01 01:00:00
2 3 2012-07-01 01:00:00
3 4 2012-07-01 01:00:00
```

After bringing the table in to kdb+ the time column is showing as foreign

```
table:.p.wrap .p.pyget`table
qdict:table[`:to_dict;`list]`
qdict
```

```
easy| 1 2 3 4
time| foreign foreign foreign
```

Let's try and bring across the underlying numeric value

qparquet - casting.md 4/4/2019

```
p)table['time'] = pd.to_numeric(table['time'])
p)print(table)
```

```
easy time
0 1 134110440000000000
1 2 13411044000000000
2 3 13411044000000000
3 4 134110440000000000
```

Casting directly to timestamp gives incorrect times

```
table:.p.wrap .p.pyget`table
qdict:table[`:to_dict;`list]`
@[qdict;`time;`timestamp$]
```

Knowing that the time epoch in Kdb+ is 2000.01.01 we can see see that python is using 1970.01.01

```
timediff: 2042.07.01D01-2012.07.01D01
2000.01.01D00-timediff
```

```
1970.01.01D00:00:00.00000000
```

Using the difference in epochs we can now perform the correct casting

```
epochOffsetNS:`long$2000.01.01D-1970.01.01D
@[`qdict;`time;{`timestamp$x-epochOffsetNS}]
flip qdict
```

qparquet - casting.md 4/4/2019

- 3 2012.07.01D01:00:00.000000000
- 4 2012.07.01D01:00:00.000000000