FINISHED

```
%pyspark
from pandas import Series, DataFrame
import numpy as np, pandas as pd
df = DataFrame([[1.4,np.nan],[7.1,-4.5],
                [np.nan,np.nan],[0.75,-1.3]],
                index=['a','b','c','d'],
                columns=['one','two'])
df
```

```
    one  two
a  1.40  NaN
b  7.10 -4.5
c   NaN  NaN
d  0.75 -1.3
```

FINISHED

```
%pyspark
df.sum()
```

```
one    9.25
two   -5.80
dtype: float64
```

FINISHED

```
%pyspark
df.mean(axis=1,skipna=False)
```

```
a      NaN
b    1.300
c      NaN
d   -0.275
dtype: float64
```

FINISHED

```
%pyspark
df.idxmax()
```

```
one    b
two    d
dtype: object
```

FINISHED

```
%pyspark
df.describe()
```

```
            one       two
count  3.000000  2.000000
mean   3.083333 -2.900000
std    3.493685  2.262742
min    0.750000 -4.500000
25%         NaN       NaN
50%         NaN       NaN
75%         NaN       NaN
max    7.100000 -1.300000
```

FINISHED

```
%pyspark
obj = Series(['a','a','b','c'] * 4)
obj
obj.describe()
```

```
count     16
unique     3
top        a
freq       8
dtype: object
```

FINISHED

```
%pyspark
from pandas_datareader import data as web
all_data = {}
for ticker in ['AAPL','IBM','MSFT','GOOG']:
  all_data[ticker] = web.get_data_yahoo(ticker)
all_data['GOOG']
```

```
2017-04-03  838.549988
2017-04-04  834.570007
2017-04-05  831.409973
2017-04-06  827.880005
2017-04-07  824.669983
2017-04-10  824.729980
2017-04-11  823.349976
2017-04-12  824.320007
2017-04-13  823.559998
2017-04-17  837.169983
2017-04-18  836.820007
2017-04-19  838.210022
2017-04-20  841.650024
2017-04-21  843.190002
[1838 rows x 6 columns]
```

FINISHED

```
%pyspark
price = DataFrame({tic: data['Adj Close']
    for tic, data in all_data.items()})
volume = DataFrame({tic: data['Volume']
    for tic, data in all_data.items()})
```

FINISHED

```
%pyspark
returns = price.pct_change()
returns.tail()
```

```
                AAPL      GOOG       IBM      MSFT
Date
2017-04-17  0.005530  0.016526  0.009261  0.008160
2017-04-18 -0.004442 -0.000418 -0.006137 -0.001375
2017-04-19 -0.003683  0.001661 -0.049162 -0.005352
2017-04-20  0.012511  0.004104  0.003773  0.007073
2017-04-21 -0.001193  0.001830 -0.011830  0.013740
```

FINISHED

```
%pyspark
print(returns.MSFT.corr(returns.GOOG))
print(returns.MSFT.cov(returns.GOOG))
```

```
0.470856307085
0.000104557652802
```

FINISHED

```
%pyspark
print(returns.corr())
print(returns.cov())
```

```
          AAPL      GOOG       IBM      MSFT
AAPL  1.000000  0.409598  0.381316  0.389620
GOOG  0.409598  1.000000  0.400583  0.470856
IBM   0.381316  0.400583  1.000000  0.493277
MSFT  0.389620  0.470856  0.493277  1.000000
          AAPL      GOOG       IBM      MSFT
AAPL  0.000265  0.000103  0.000074  0.000091
GOOG  0.000103  0.000240  0.000074  0.000105
IBM   0.000074  0.000074  0.000143  0.000085
MSFT  0.000091  0.000105  0.000085  0.000206
```

FINISHED

```
%pyspark
print(returns.corrwith(returns.GOOG))
print("\n")
print(returns.corrwith(volume))
```

```
AAPL    0.409598
GOOG    1.000000
IBM     0.400583
MSFT    0.470856
dtype: float64
AAPL   -0.072783
GOOG   -0.009838
IBM    -0.204049
MSFT   -0.089592
dtype: float64
```

READY