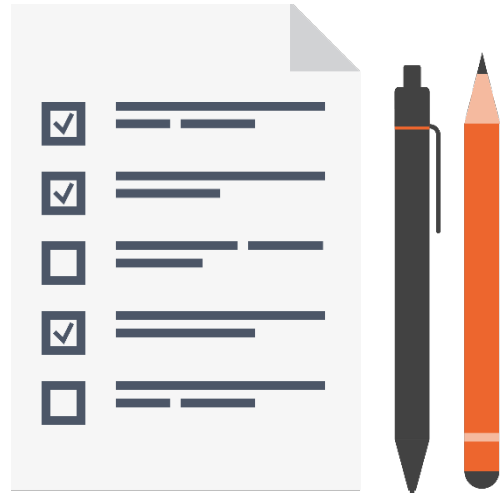


Using Pig with Real Data

Overview



Loading & Storage

Diagnostic Operators

Real World Demo

Load

```
var = LOAD '/tmp/datafile.csv' USING PigStorage(',')  
      (name:chararray, price:int);
```



Load Using PigStorage

```
var = LOAD '/user/hue/NDX-100.csv' USING  
PigStorage(',') AS (name:chararray, price:int);
```

Store Using PigStorage

```
STORE var INTO 'filename' USING PigStorage(',');
```

Other Options

```
PigStorage(',')  
field1,field2,field3
```

```
PigStorage(':')  
field1:field2:field3
```

```
PigStorage('*')  
field1*field2*field3
```

Dump

```
var = LOAD '/tmp/datafile.csv' USING PigStorage(',')  
AS (date, open, high);
```

```
DUMP var;
```

Describe

```
var = LOAD '/tmp/datafile.csv' AS
```

```
(name:chararray, price:int);
```

```
DESCRIBE var;
```


Explain

```
var = LOAD '/tmp/datafile.csv' AS  
  
(name:chararray, price:int);  
  
EXPLAIN var;
```

Illustrate

```
var = LOAD '/tmp/datafile.csv' AS
```

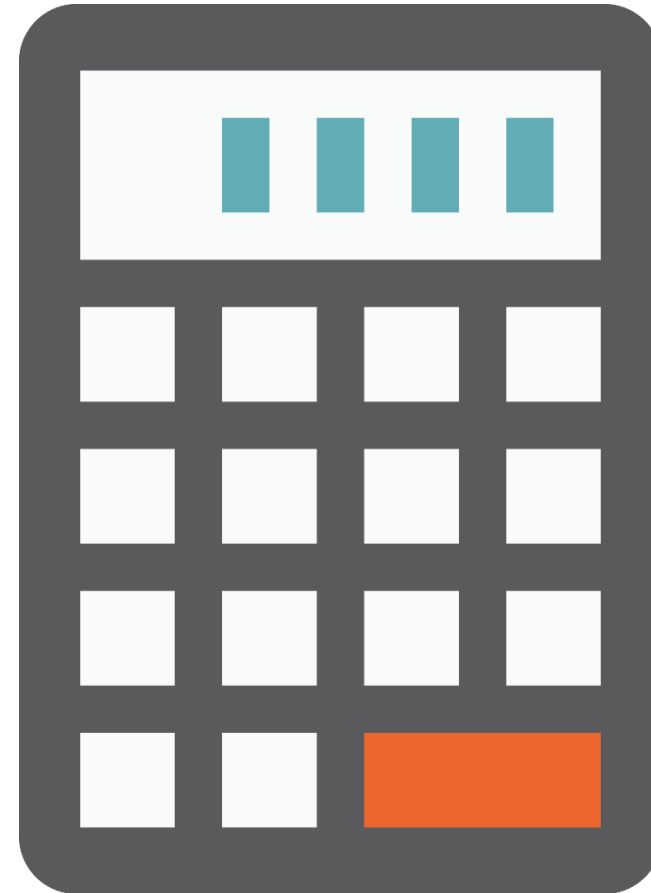
```
(name:chararray, price:int);
```

```
ILLUSTRATE var;
```

Stock Data

Historical Stock Data

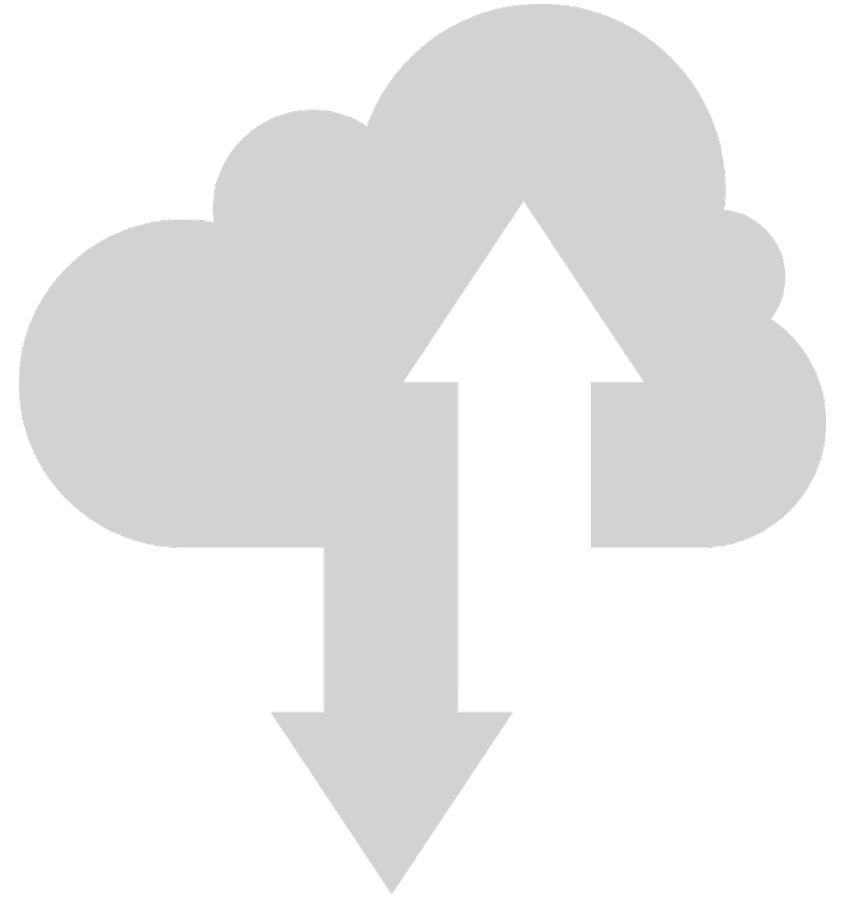
www.yahoo.com/



Weather Data

Historical Weather Data

<http://www.ncdc.noaa.gov>



Load Data

Use PigStorage() to Load data

Set Data Types



Format Dates

Problems with data
ToString()

```
OBS BELVEDERE TOWER NY US,20080325,89,-6
OBS BELVEDERE TOWER NY US,20080326,161,61
OBS BELVEDERE TOWER NY US,20080327,94,67
OBS BELVEDERE TOWER NY US,20080328,89,33
OBS BELVEDERE TOWER NY US,20080329,83,6
OBS BELVEDERE TOWER NY US,20080330,94,-22
OBS BELVEDERE TOWER NY US,20080331,139,28
OBS BELVEDERE TOWER NY US,20080401,194,122
```

Compare

Join stock prices and weather data

Store results in CSV



Real World Demo

Weather Data

Stock Data



Summary

That's all
folks.



Load, Dump, PigStorage

Describe, Explain, Illustrate

Walkthrough Real World Example