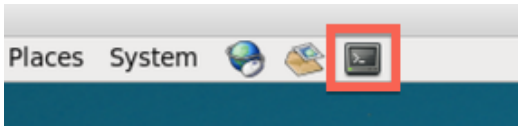
[Download a PDF of this hands-on tutorial]

ExploringSensorData.pdf

By the end of this activity, you will be able to:

1. Identify the major components in semi-structured data from a weather station
2. Create plots of weather station data

Step 1. Open a terminal shell. Open a terminal shell by clicking on the square black box on the top left of the screen.



Change into the sensor directory:

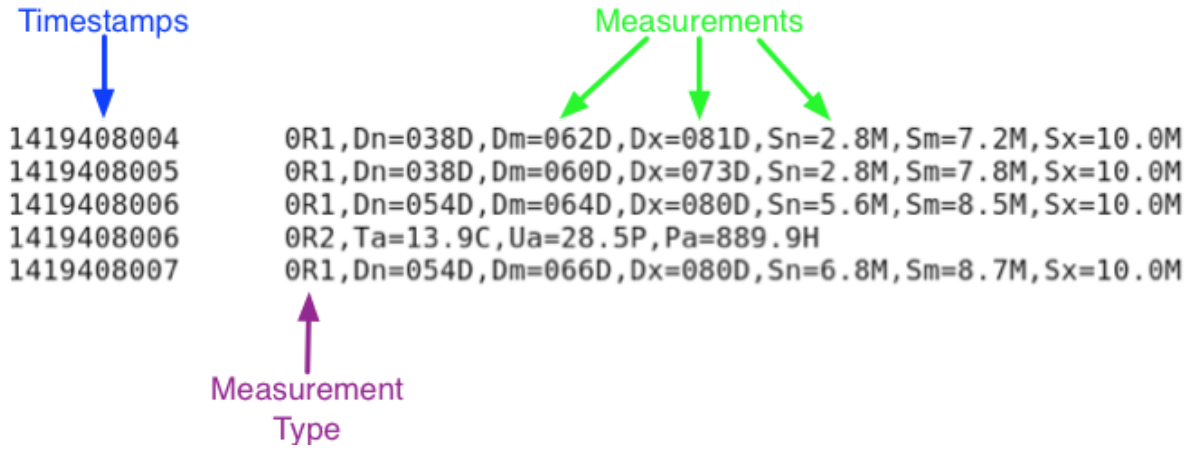
```
cd Downloads/big-data-2/sensor
```

Run ls to see the sensor data files and scripts:

```
[cloudera@quickstart sensor]$ ls
plot-data.py    stream-plot-data.py  wxt-520-format.txt
stream-data.py  wx-data.txt
```

Step 2. View weather station data. Twenty-four hours of weather station data is in the file *wx-data.txt*. View this data by running *more wx-data.txt*:

```
more wx-data.txt
```



Each line in the file contains a timestamp in the first column and a set of measurements in the second column. These column are separated by a tab. The measurements are also a set of columns separated by commas.

The timestamp is the number of seconds since January 1, 1970. By looking at the successive timestamps in the file, we can see that the measurements arrive about every second.

The measurements are prefixed with *R0, R1, R2*, etc. Each R-value holds a different set of measurements. For example, *R1* has *Dn, Dm, Dx*, etc. Additionally, the different types are measured at different frequencies. R1 is measured every second, but R0 and R2 are much less frequent.

Step 3. Look at key for measurements. We can learn what the measurement fields mean by looking at the measurements key file. Open a new terminal and *cd* to *Downloads/big-data-2/sensor*. Run more *wxt-520-format.txt*:

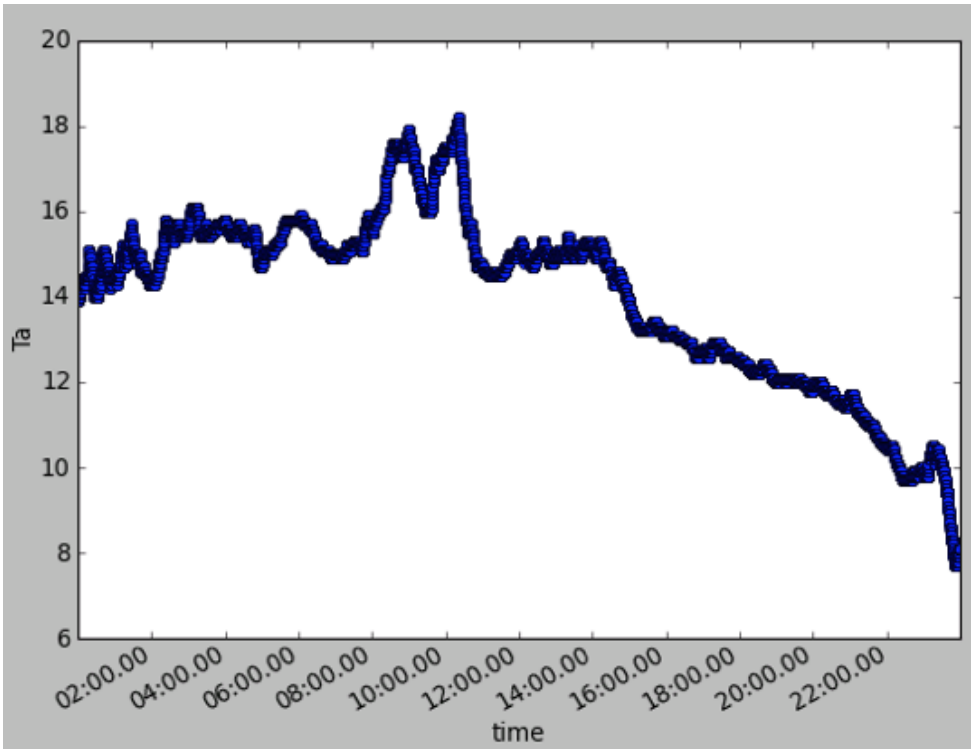
```
1  more wxt-520-format.txt
```

```
Sn      Wind speed minimum m/s, km/h, mph, knots #,M, K, S, N
Sm      Wind speed average m/s, km/h, mph, knots #,M, K, S, N
Sx      Wind speed maximum m/s, km/h, mph, knots #,M, K, S, N
Dn      Wind direction minimum deg #, D
Dm      Wind direction average deg #, D
Dx      Wind direction maximum deg #, D
Pa      Air pressure hPa, Pa, bar, mmHg, inHg #, H, P, B, M, I
Ta      Air temperature °C, °F #, C, F
Tp      Internal temperature °C, °F #, C, F
Ua      Relative humidity %RH #, P
Rc      Rain accumulation mm, in #, M, I
Rd      Rain duration s #, S
Ri      Rain intensity mm/h, in/h #, M, I
Rp      Rain peak intensity mm/h, in/h #, M, I
Hc      Hail accumulation hits/cm2, hits/in2, hits #, M, I, H
Hd      Hail duration s #, S
Hi      Hail intensity hits/cm2h, hits/in2h, hits/ h #, M, I, H
Hp      Hail peak intensity hits/cm2h, hits/in2h, hits/ h #, M, I, H
Th      Heating temperature °C, °F #, C, F
Vh      Heating voltage V #, N, V, W, F2
Vs      Supply voltage V V
Vr      3.5 V ref. voltage V V
```

This file provides a list of what each field means. For example, *Ta* is the Air temperature, and *Dn* is the Wind direction minimum.

Step 4. Plot data. Run *plot-data.py wx-data.txt Ta* to plot the air temperature:

```
1  ./plot-data.py wx-data.txt Ta
```



Note: If you see the following error**:** ImportError: No module named matplotlib.pyplot, your python environment is being incorrectly pointed to. To find the Math Plot Library, edit the first line to

```
1  #!/usr/bin/env python
```

You can plot other measurements in the file by using the field name as the last argument. For example, use Dn to plot the maximum wind direction:

```
1  ./plot-data.py wx-data.txt Dn
```

Mark as completed