

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

1. View semi-structured data streaming in real-time from a weather station
2. Create plots of streaming 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:

```
1 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 streaming weather station data.** Run `stream-data.py` to see streaming data from the weather station:

```
1 ./stream-data.py
```

The measurements are appearing as they are produced by the weather station. By looking at the timestamp, we can see that they arrive about every second. Additionally, different measurement types are produced at different frequencies. For example R1 is measured every second, but R2 is less frequent.

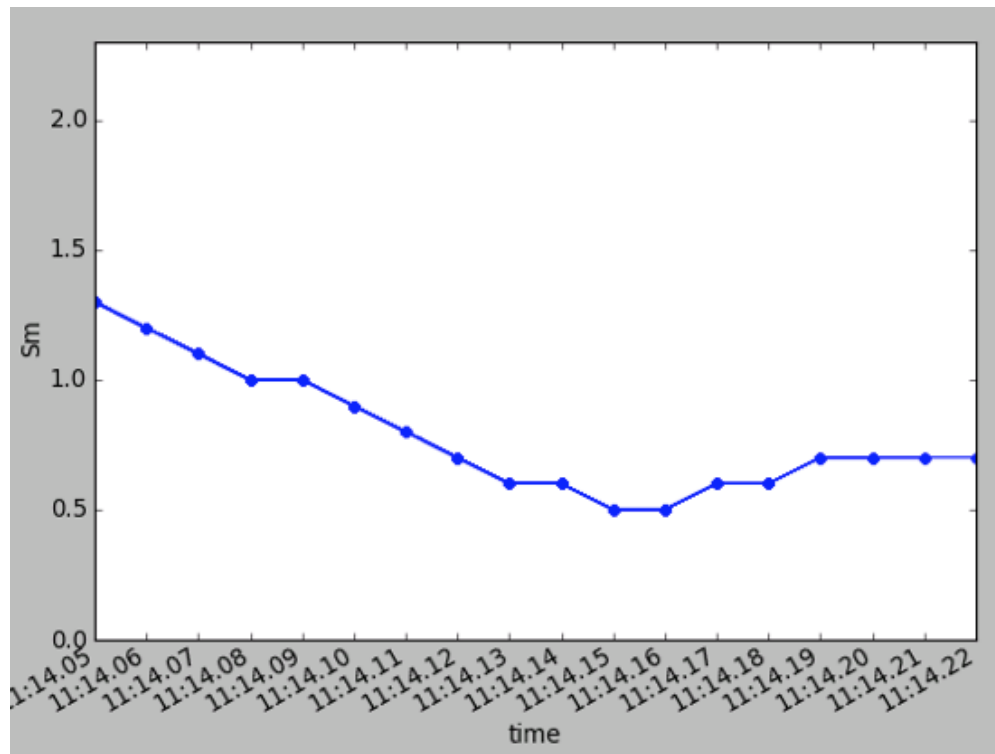
**Step 3. View key for measurements.** Run more wxt-520-format.txt to see the key for the measurement fields.

```
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
```

**Step 4. Create plot of streaming data.** We can plot the streaming data by running stream-plot-data.py:

```
1 ./stream-plot-data.py Sm
```



This shows a plot of the average wind speed ( $S_m$ ), and it is updated every time a new measurement is produced. Notice that the plot is updated every second. We can create plots for other measurements by specifying the measurement in the last argument to `stream-plot-data.py`. For example, we can plot air temperature ( $T_a$ ) by running:

```
1 ./stream-plot-data.py Ta
```

This plot is updated less frequently than the first plot since air temperature measurements are produced less frequently.

Mark as completed



