**Date:** May 10, 2021

**Subject:** Discussion on InfluxDB

**Attendees:**

* **MDS Team:** Nathan Smith, Mitch Harris, Ryan Koenig
* **UDL Team:** Jiachen Wei,

**Agenda:**

**1 – SkySpark Historical Questions**

Understanding available data for CEC and BRDF (we’re not sure we’re finding all the data on InfluxDB). Understanding the schema that has been setup (points vs readings) and a general discussion on tags and field in InfluxDB.

Current CEC/BRDF SkySpark data in InfluxDB is limited, need to update Telegraf file to allow parsing of additional data. Nate/Ryan to assist with this – Jiachen to provide details.

Tags relate to more common queries vs fields.

From the schematic: https://github.com/UBC-UrbanDataLab/SkySpark\_data

Readings consist of the limited set of tags and are updated continuously whereas points are updated daily and can have ~150 tags.

**2 – SkySpark Streaming Questions**

How will InfluxDB ingest the streaming SkySpark data (Telegraf?)

Yes, currently being setup with the correct JSON parsing (Nate/Ryan to support).

**3 – InfluxDB Processing**

How has InfluxDB processing been done to date? From the examples looked at, most involve reading from a bucket, processing, and writing to another bucket but there don’t appear to be any processed buckets – potentially any processing is being written directly to the SkySpark bucket itself? This would be relevant to where anomalous flags could go.

Likely setup anomalous readings in another bucket, TBD.

General thoughts streaming processing from InfluxDB (Flux Tasks, Telegraf). Note that processing in Influx v2.0 is different from v1.\*.

Options include: using anomaly detection with Telegraf before data are written to InfluxDB, or using Flux Tasks.

New user with write privileges to the MDS2021 bucket

Username: MDS2021

Pwd: \*\*\*\*\*\*\*\*

Token (can also be found on the GUI)

\*\*\*\*\*\*\*\*

Install open-source Influx 2.0 and Telegraf

(UDL uses Debian on Compute Canada, so my commands below are Linux commands for reference)

Download: <https://portal.influxdata.com/downloads/>

[Get Started with InfluxDB 2.0](https://docs.influxdata.com/influxdb/v2.0/get-started/?t=Linux)

### [Set up InfluxDB through the UI](https://docs.influxdata.com/influxdb/v2.0/get-started/?t=Windows#set-up-influxdb-through-the-ui)

# [Get started with Telegraf](https://docs.influxdata.com/telegraf/v1.17/introduction/getting-started/)

More tutorials

[Automatically configure Telegraf (on InfluxDB 2.0 GUI)](https://docs.influxdata.com/influxdb/v2.0/write-data/no-code/use-telegraf/auto-config/#create-a-telegraf-configuration)

You can also check out [Telegraf Playground](https://rootnroll.com/d/telegraf/)

For the plugins that we cannot auto-configure, we will manually edit the conf file.

1. Store authentication token as an [environment variable](https://wiki.debian.org/EnvironmentVariables)

export **INFLUX\_TOKEN**=TokenCopiedFromUI #no space around “=”

echo $INFLUX\_TOKEN

1. Generate a configuration file

cd telegraf

telegraf config > telegraf.conf

sudo nano /etc/telegraf/telegraf.conf

1. Edit the sections for our selected parser and output plugin.

<https://github.com/influxdata/telegraf>

* Input plugin: you might not need this (UDL uses [http\_listener\_v2](https://github.com/influxdata/telegraf/blob/master/plugins/inputs/http_listener_v2) ; you can see the full list on input plugins [here](https://github.com/influxdata/telegraf#input-plugins))
* Parser: [JSON](https://github.com/influxdata/telegraf/blob/master/plugins/parsers/json)

I believe we need to specify these variables in the conf file

files = [“a/sample/json/file”]

data\_format, json\_time\_key, json\_time\_format, tag\_keys, and maybe more

The sample json files are in [this OneDrive folder](https://ubcca-my.sharepoint.com/:f:/r/personal/michael_kennedy_ubc_ca/Documents/Urban%20Data%20Lab/7.0%20Active%20Research%20Projects/Project%20008%20-%20MDS%20Capstone%202021/Sample_JSON_Data?csf=1&web=1&e=RNiE7K)

* Output plugin: [influxdb\_v2](https://github.com/influxdata/telegraf/blob/master/plugins/outputs/influxdb_v2)

Comment [[outputs.influxdb]]

Uncomment [[outputs.influxdb\_v2]]

urls = ["<http://locahost:8086>"]

token = "$INFLUX\_TOKEN"

organization = "UBC"

bucket = "MDS2021"

1. Save and exit the conf file. [Start the Telegraf service.](https://docs.influxdata.com/telegraf/v1.17/introduction/getting-started/#start-telegraf-service)

**Schedule:**

* Currently working on Week 2 – generally going well.

**Recent Progress:**

* Proposal complete
* Working on Python streaming simulation, anomaly detection algorithm review, EDA

**Next Steps:**

* Additional task of helping parse data using Telegraf which will be helpful for determining if anomaly detection should be done with the Telegraf step
* Finish Week 2 and kick-off Week 3 which includes implementing anomaly detection