

Data Science Infrastructures – Exercise 04 (DSI E04 ST 24)

16/05/2024 / Philipp Wieder / philipp.wieder@gwdg.de

General Prerequisites

To fulfil this assignment, you need an InfluxDB Stack running on your machine. Luckily, this is an easy task using the so-called *sandbox*. Other prerequisites, which most of you should have fulfilled already, are Docker and git (for Docker, check Exercise 02, for git check <https://git-scm.com/book/en/v2/Getting-Started-Installing-Git>). To install the InfluxDB sandbox locally, just use the following command

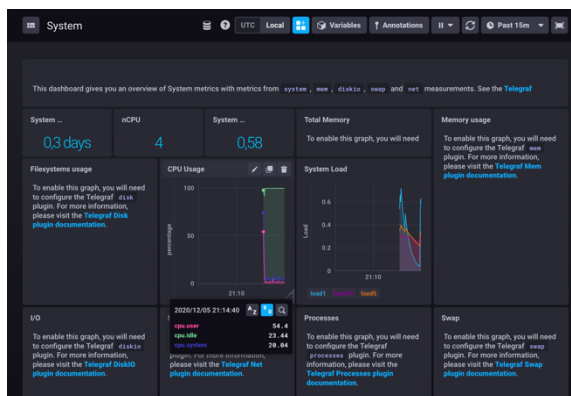
```
git clone https://github.com/influxdata/sandbox.git
```

This install a directory `/sandbox` in your current directory. Switch to `/sandbox` (in case you forgot: you can use the `cd` command). Then start the InfluxDB stack with

```
./sandbox up1
```

It takes a while, but then directly opens two browser windows showing Chronograf on `localhost:8888`, which is the UI for InfluxDB, and the sandbox documentation on `localhost:3010`. Now you are prepared.²

Assignment I – Setup & Explore



Points: 2

Going to `localhost:8888` you will see the Chronograf standard UI. Choose the *Configuration* menu item and click on the only existing connection *My InfluxDB (Default)*. This opens a *Connection Configuration* dialogue, where you click *Update Connection*. Select *System* and then click *Create 1 Dashboard*. Click *Continue* and then *Finish*. Go to dashboards and select *System*. This will bring you to a dashboard showing your system status (with some missing parts that are not yet configured). Explore the *CPU Usage* and *System Load* parts. Also have a look at the documentation on `localhost:3010`.

¹ It might happen that you get an error message during the start of chronograf component: *"Error response from daemon: invalid mount config for type "bind": bind source path does not exist: <your_path_to_sandbox>/sandbox/chronograf/data"*. In case this happens, just create the directory `./chronograf/data` and things should work.

² The sandbox does not contain the latest-greatest InfluxDB version, but it is the easiest way to install it. And it is sufficient to get a good understanding how it works and what capabilities it offers.

Assignment

In the System dashboard, we only see a few time series displayed, but many Telegraf plug-ins are not enabled. This we are going to change now.

Enable the Filesystem usage, Memory usage, and I/O plug-ins. You can also try others, but first check whether your operating system is supported. To do this, you have to edit the Telegraf configuration file. You find it at `./telegraf/telegraf.conf` in your sandbox folder. You need to add the following:

```
[[inputs.disk]]
## Ignore mount points by filesystem type.

ignore_fs = ["tmpfs", "devtmpfs", "devfs", "iso9660", "overlay", "aufs",
"squashfs"]
# Read metrics about disk IO by device
[[inputs.diskio]]
# Read metrics about memory usage
[[inputs.mem]]
```

Each `inputs.xxx` line enables the respective plug-in so that you can see the data in the dashboard. Now you need to restart the sandbox: `./sandbox restart`. Go back to `localhost:8888` and check the result (it takes some time to start the web server again, so be patient). As a result, you have quite a good overview of your system performance³. Submit a screenshot as your solution.

Assignment II – Querying the InfluxDB

Points: 2

In addition to configuring the existing dashboard with system-specific data, you can also create a new one. This is what we do in this assignment.

Assignment

- Use the *Dashboard* menu item on the to check the existing dashboards. Create a new one and name it "Memory".
- Add data to the dashboard. By clicking "Add Data", you will enter a dialog where you can specify an InfluxQL query and directly see the results of the query. The correctness of the query is also checked. By clicking on "telegraf.autogen" you see the available measurements.

Now create two widgets displaying the following data:

- A memory monitoring. Search *Measurements & Tags* for `mem` select the `used` field. Name the widget *Used Memory*. Display the data of the last hour. Adjust the

³ The data is persistently stored at `./influxdb/data/data/telegraf/_series`.

visualization with a proper y-axis title, choose the right scale (called Y-Value's Format), and make sure that the legend is displayed.

2. Add a similar widget for the memory available. Name it *Available Memory* and also format it according to the rules above.

Submit the two queries and a screenshot of the dashboard as a result of this assignment.