

Contents

[Description 2](#_Toc514781740)

[How to Do It? 3](#_Toc514781741)

[Installing DNIF 3](#_Toc514781742)

[Identifying the dataset and its source 4](#_Toc514781743)

[Understanding DNIF 5](#_Toc514781744)

[Capturing the Dataset 6](#_Toc514781745)

[Uploading the data 8](#_Toc514781746)

[Creating Package 8](#_Toc514781747)

[Creating Widget(s) 8](#_Toc514781748)

[Creating Dashboard 8](#_Toc514781749)

[Configuring SMTP for DNIF 8](#_Toc514781750)

[Creating Alert(s) 8](#_Toc514781751)

# Description

**Process 1: Static Data in DNIF**

The process is divided into 4 stages of small description. This will help us further to understand how we are solving problems in each stage or what is necessary for that particular stage to be completed and pass on to the next stage.

*Stage 1:*

* Select & understand data-set from a domain of interest.
* This is the initial step, which of-course sets the entire purpose of what kind of data you are looking to analyze and make visuals on. It is an important step as your quality of work depends entirely on this. (There is no limitation to the number of data sources, it will only be limited by imagination and lack of skills in further stages – meaning multiple datasets could be merged in to one single source using logic and code & then uploaded on DNIF)

*Stage 2:*

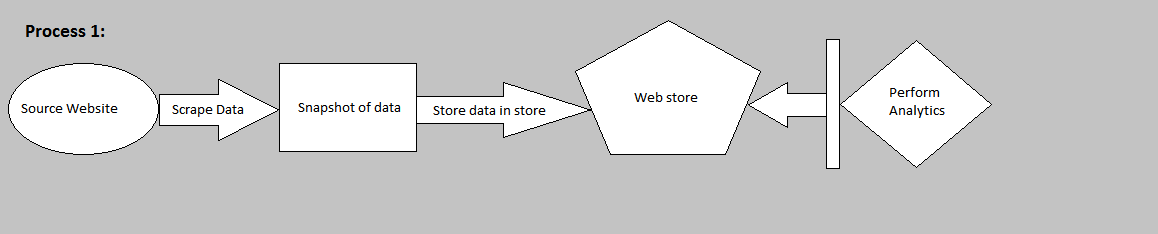
* Understand DNIF platform and its capabilities limited to project scope.
* This step was introduced so that, one is beforehand ready and understands what the platform is, how it works, what are its capabilities, its features, the query language, deployment models, limitations, etc.

*Stage 3:*

* Get the static data-set inside DNIF platform
* This stage is where the real game begins. Look out for the format of the file one is uploading, how it is being uploaded, and other requisites.

*Stage 4:*

* Perform analytics, create dashboards, & build alerts.
* Once your data is inside DNIF by uploading through the event store, all the skills and understanding acquired in stage 2 could be put into action.



*Fig 1: A Squeezed Perspective*

# How to Do It?

Let us go in steps in order to achieve the outcome desired. These steps are descriptive in details and do not directly highlight which stage they accord to. Let us dive in and perform these sequentially.

**STEP 1:**

## Installing DNIF

* To do this, first “Sign-Up” with DNIF and read the pre-requisites guide here:

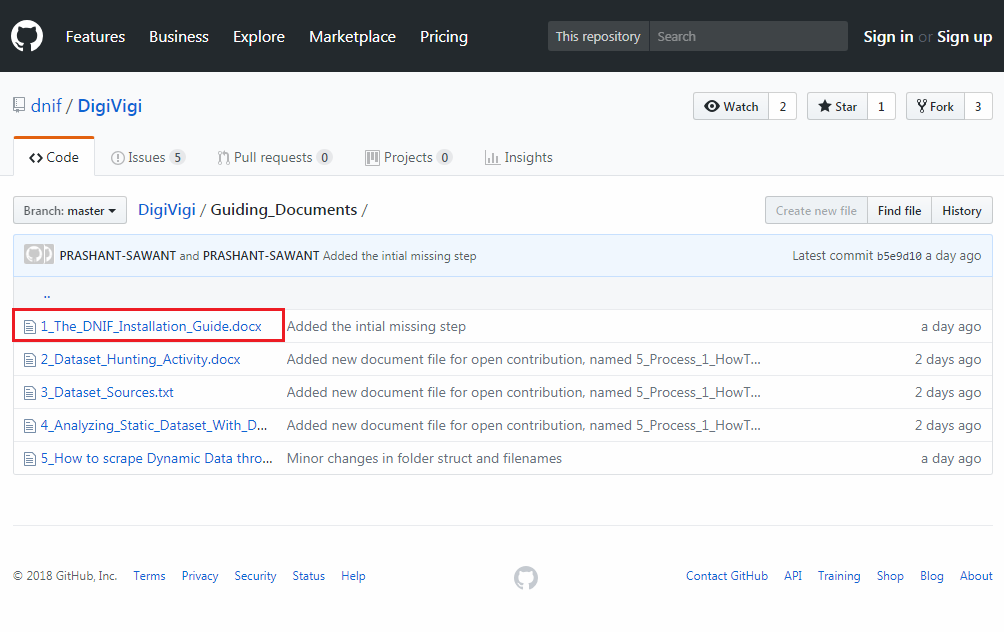
Sign-up: <https://dnif.it/signup.html>

Pre-requisites: <https://dnif.it/docs/guides/getting-started/prerequisites.html>

* Then follow the steps mentioned in “1\_The\_DNIF\_Installation\_Guide.docx” which is present in the repository. Here’s a link to it: <https://github.com/dnif/DigiVigi/tree/master/Guiding_Documents>
* There are videos and other documents as well which are provided by the Organization which can guide, but it does not cover the details in it which we require. Here’s a link to it:

<https://dnif.it/docs/guides/getting-started/installing-dnif.html>

**NOTE:** In our case we have done the setup on a Virtual Machine. One can have a separate server like machine with a dedicated Ubuntu/ CentOS.



*Fig 2: DNIF Installation Pointer*

**STEP 2:**

## Identifying the dataset and its source

* Here’s a guide of how we did it for starters. Refer following documents:

1. Activity initialization:

<https://github.com/dnif/DigiVigi/blob/master/Guiding_Documents/2_Dataset_Hunting_Activity.docx>

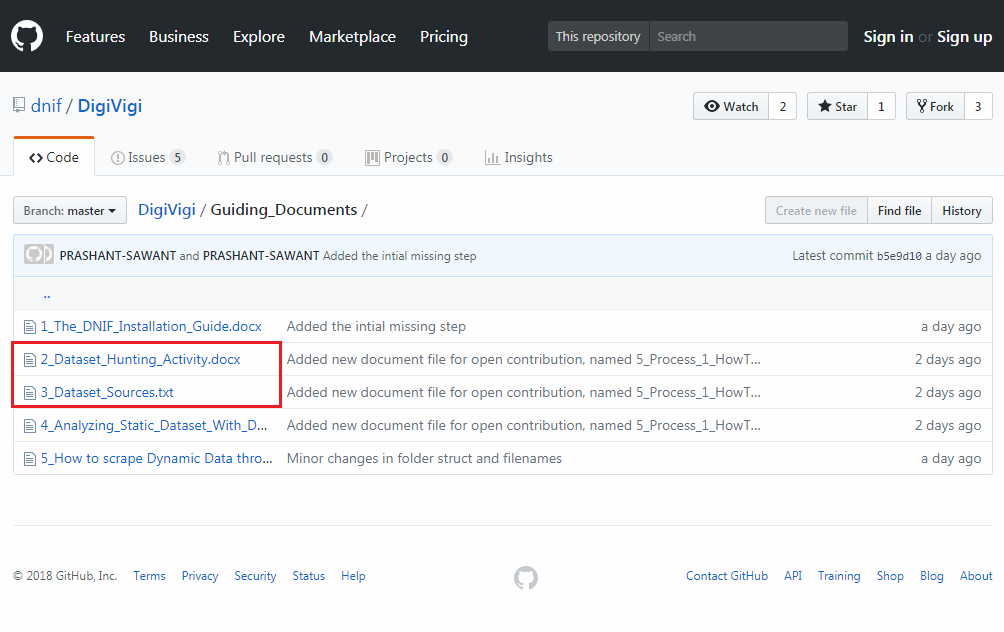
1. Jotting down dataset sources

<https://github.com/dnif/DigiVigi/blob/master/Guiding_Documents/3_Dataset_Sources.txt>

* Although it is not a must that our dataset sources might be of a good use to most analysts, since the purposes might differ.
* So in this current documented example, we’ve chosen a structured and static dataset from the source:

<https://www.webiron.com/abuse_feed/>

* Webiron is not a static source. In fact it posts newer data everyday on its website link mentioned above. Only in our case we have scrapped data from the web page for a course of 3-5 days just for demonstration purpose.



*Fig 3: Dataset Repoistory*

**STEP 3:**

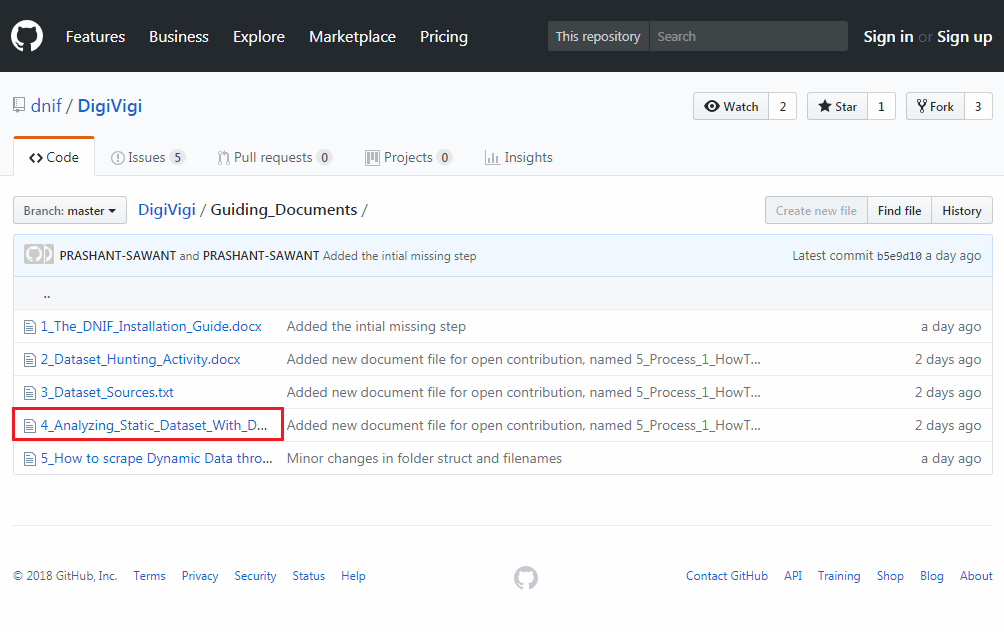
## Understanding DNIF

* This one here is a self-study step. It won’t be much of a tutorial based way through this document. But there’s no need worry, as DNIFs query language is very simple to learn if one is well versed with SQL. Yes, heard it right. They have this entire documentation of the query commands with examples.
* Head over here for getting an idea about it:

<https://dnif.it/docs/>

* The link above has everything in it for one to understand DNIF on a practical basis. From query language to tutorials and other documentations, most of it is present here originally.
* Also one of our contributors to “DNIF Open Source Project” has already given a glimpse of data analysis, widgets, dashboards and packages under the repository.

<https://github.com/dnif/DigiVigi/blob/master/Guiding_Documents/4_Analyzing_Static_Dataset_With_DNIF.docx>



*Fig 4: Analysis Documentation*

**STEP 4:**

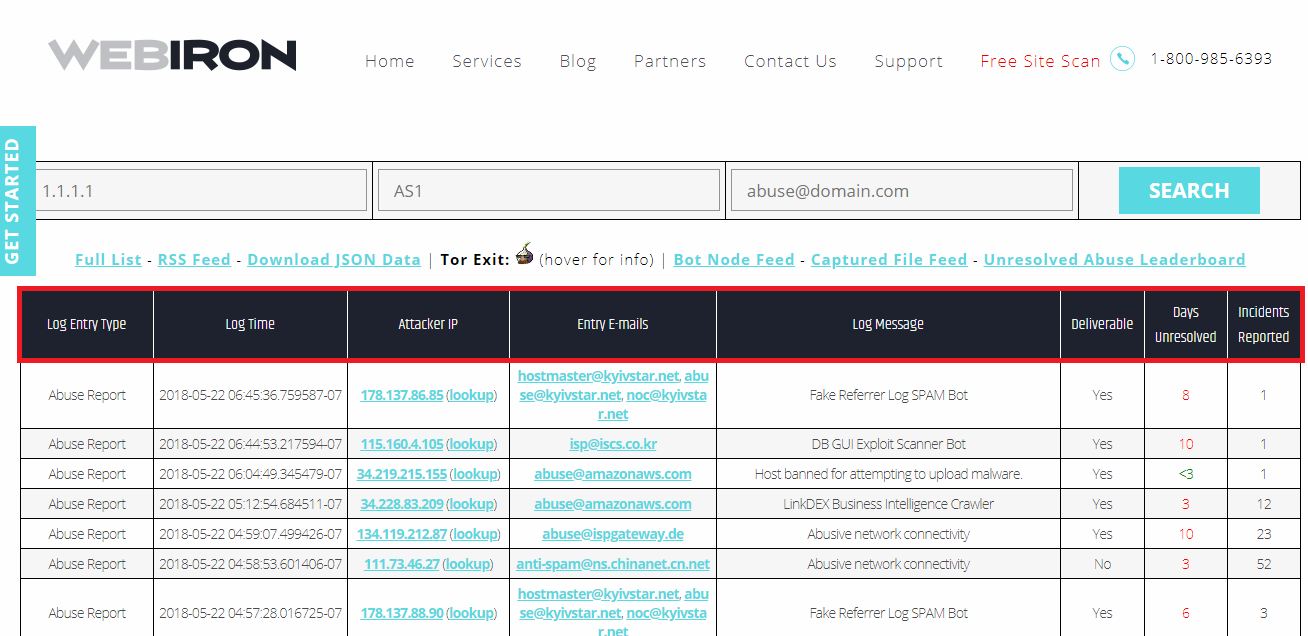
## Capturing the Dataset

Next thing is where we have written a python script which uses “Web Scrapping” – a technique used to scrape data off the HTML page.

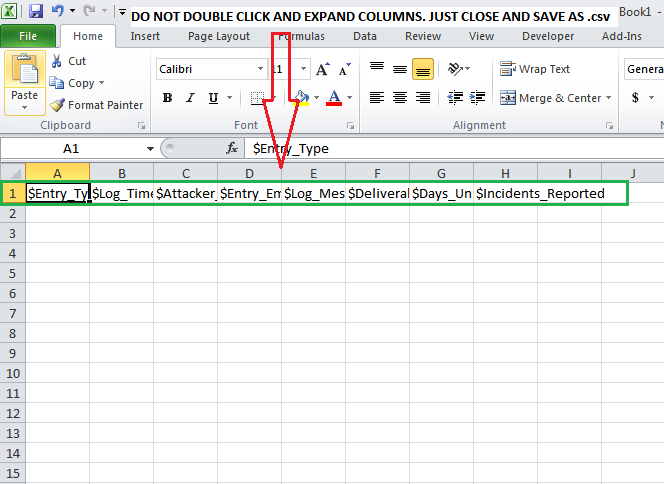
* Let us flash back a little
  + We have DNIF Installed with a connection to the container established.
  + We have our dataset source
  + Let us create a “csv” file named ‘HistoricalData.csv’ with headers predefined. This is a one-time manual thing after that every time the script is run; the data will just get appended.
  + We are being very careful with headers here since DNIF has a very convenient way of fetching records when queried. Also these headers were identified from the dataset source link.

1. Open Excel
2. Type in the headers with '$’ sign being the first character
3. Do not expand the columns, since this file is going to be saved as ‘HistoricalData.csv’
4. To save, just close and save it as ‘Comma Delimited’ file type.

* + Here’s a glimpse of how the file was created:

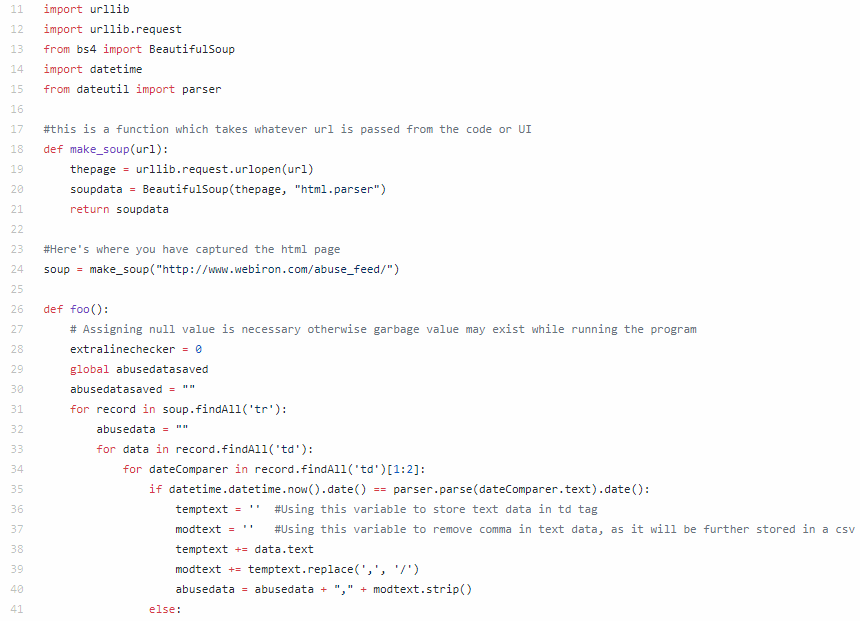


*Fig 5: Webiron Abuse Feed Header Highlight*



*Fig 6: Making that storage file*

* + Next we’ll scrape that data, using this script:



*Fig 7: Code Glimpse*

* Above is not the entire code. It is present on the repository named “PS\_KeepEveryDayData.py”

<https://github.com/dnif/DigiVigi/blob/master/Process_1/PS_KeepEverydayData.py>

**NOTE:** If using a different website source with numerically different column numbers, please modify the code and CSV storage file accordingly.

**STEP 5:**

## Uploading the data

**STEP 6:**

## Creating Package

**STEP 7:**

## Creating Widget(s)

**STEP 8:**

## Creating Dashboard

**STEP 9:**

## Configuring SMTP for DNIF

**STEP 10:**

## Creating Alert(s)