# rrd - Remove data from RRDTool - Stack Overflow

http://stackoverflow.com/questions/10298484/remove-data-from-rrdtool

I have several graphs created by RRDTool that collected bad data during a time period of a couple hours.

How can I remove the data from the RRD's during that time period so that it no longer displays?

Best method I found to do this...

1. Use [RRDTool Dump](http://oss.oetiker.ch/rrdtool/doc/rrddump.en.html) to export RRD files to XML.
2. Open the XML file, find and edit the bad data.
3. Restore the RRD file using [RRDTool Restore](http://oss.oetiker.ch/rrdtool/doc/rrdrestore.en.html) .

If you want to avoid writing and editing of xml file as this may takes few file IO calls(based on how much bad data you have) , you can also read entire rrd into memory using fetch and update values in-memory.

I did similar task using python + rrdtool and i ended up doing :

1. read rrd in-memory in a dictionary
2. fix values in the dictionary
3. delete existing rrd file
4. create new rrd with same name.

# O+P [rrd-users] Cleaning a messed up rrdtool data file

https://lists.oetiker.ch/pipermail/rrd-users/2005-March/009512.html

List,

I have an rrdtool database that polls my main router to the net. From

time to time there was an ugly 70Mb spike in the graph, and I didn't pay

too much attention, thinking that it was just some sort of counter

rollover or something. It was never a sufficient annoyance to bother

looking into it more closely.

This week, the line was upgraded from 2Mb to 6Mb and as I was puttering

around with my graphing tools I noticed that I was running two copies of

my polling script that was updating the rrd file. And so it makes sense

to me that that was the source of the spikes. Now I'd like to try and

clean up the historical data.

I've generated two graphs, one clamped rigidly to the maximum expected y

value, and the other with y autoconfigured:

<http://www.landgren.net/rrdtool.html>

At first I was thinking I could use something like

result=number,256000,GT,UNKN,number,IF

but then I'll clamp new legitimate values that'll be coming down the

wire. If I clamp to the new higher value I'll have spikes in the old

traffic values. I'd really like to have clean data so that I can point

and say "look, before... and after".

Therefore, is there a way of saying:

"if date before d0, then clamp to lim1 else clamp to lim2"

The second approach I considered would be to go into the rrd file and

iterate over the values, looking for peaks greater than 2Mb, and

replacing them by a linear interpolation between to the two points

either side of the peak?

The maths doesn't pose much of a problem, I'm wondering about the

mechanics of going through the datapoints in the rrd file.

Thanks for any pointers I can use to either approach (or if there's a

third...),

David

That would be a clean solution. You could do it like this:

- "rrdtool dump" the database

- Use a regexp to extract the DS values from your RRA rows

- Change them to what you like

- "rrdtool restore" the database back to its binary form

Serge.

-------------

Op de inhoud van dit e-mailbericht en de daaraan gehechte bijlagen is de inhoud van de volgende disclaimer van toepassing: <http://www.zeelandnet.nl/disclaimer.php>