Original Research Paper

RRDTool: A Round Robin Database for Network Monitoring

Sweta Dargad and Manmohan Singh

Department of Computer Science Engineering, ITM Universe, Vadodara 390510, India

Abstract: RRDtool is a circular database which stores time series data. RRDtool assumes time-variable data in intervals of certain length. An interval is named as a step and specified upon creation of an RRD file which cannot be changed afterward. Because data may not always come at just the right time, RRDtool will automatically interpolate any submitted data. This is done to fit its internal timestamps. If we compare linear databases with RRD then we can say that RRD is non-linear like a circular queue. Usually, in a Round-Robin Database (RRD), time-series data like network bandwidth, memory usage, CPU load are stored. The data is stored in a non-linear way so that system storage footprint remains constant over time. This actually avoids resource-expensive purge jobs which reduce complexity. MySQL does NOT provide any such kind of storage engine. RRDtool has also the ability to create and feed this database. RRDtool stores data; that makes it a back end tool. The RRDtool command set allows the creation of graphs that makes it a front end tool as well. Other databases just stores data and cannot create graphs. RRD combined with NET-SNMP can collect data and using scripts we can generate beautiful real-time graphs which can be used for monitoring computer networks.

Keywords: Network Monitoring, RRD Tool, NET SNMP, Computer Network, Graphing, Open-Source, Database, Network Management

Introduction

Round-Robin Database tool (RRDtool) handles time-series data like temperatures, CPU load, network bandwidth etc. The data is stored in a round-robin database which acts as a circular buffer, thus the system storage footprint remains constant over time. RRDtool is open source tool and it can be integrated into Shell scripts, Python, Ruby, Perl, Lua or Tcl applications easily. RRD tool assumes time-variable data in intervals of a certain length. This interval, usually named step, is specified upon creation of an RRD file and cannot be changed later.

RRDTool interpolates data to get the time steps. PDP stands for Primary Data Point. It is the value of specific step which is interpolated. Consolidation Function CF is used to plot Multiple PDPs. This plotting of PDPs forms a Consolidated Data Point (CDP). Typical consolidation functions are average, minimum, maximum (Harrington *et al*., 2002). Round Robin Archive (RRA) stores the Consolidated Data Points (CDP).

Total time covered by RRA:

Time Covered = (#CDPs Stored)_(#PDPs per CDP)_Steps

In a Round-Robin Database (RRD) usually time-series data like network bandwidth, temperatures, CPU load etc. is stored. To avoids resource expensive purge jobs and reduces complexity data is stored to form a constant footprint of data. MySQL doesn't provide this kind of storage facility. Although some people were thinking about it and some prototypes exists. RRDTool re-samples the data and store this value. Network Monitoring tools gather data after some specific time. But the data is not available all the time because of some Query lost or Device Down. Data does not arrive exactly in that interval. It takes actual arrival time under curve and creates new points in time between this time intervals and make a curve beneath the real curve. Thus (http://www.cotse.com/tools/netman.html Networktrafficmonitoring) network traffic latency will be same but the numbers shown may differ.
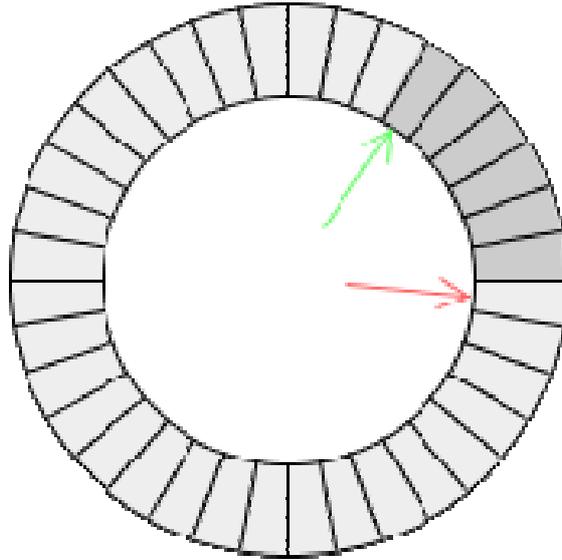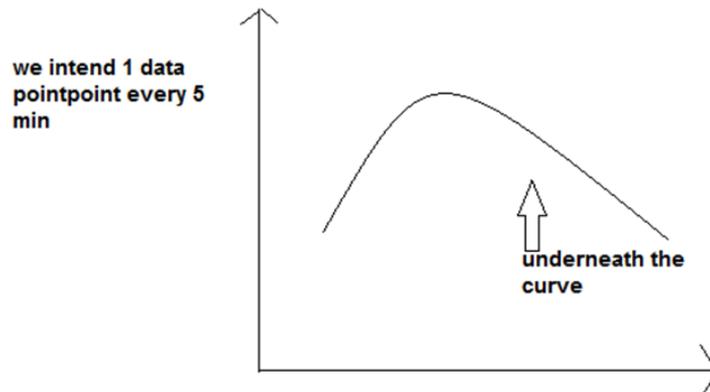
Fig. 1. RRD's circular data structure

Fig. 2. RRD's data curve

Acquisition, Consolidation and Archival

In Fig. 1 we have shown the structure of RRDtool database. The figure clearly shows that the structure is like a circular queue where data is updated as time passes. The data while the man pages talk of command line switches you have to set in order to make RRDtool work it is important to note that RRDtool can be remotely controlled through a set of pipes. This saves a considerable amount of startup time when we plan to make RRDtool do a lot of things quickly. We need to check the section on Remote Control further down.

Data Acquisition

When monitoring the state of a system, it is convenient to have the data available at a constant time interval. Unfortunately, we may not always be able to fetch data exactly at the time we want to.

RRDtool can update the log file when user specifies. RRD Tool will interpolate the data from the Data Source and write it in the log.

Consolidation Function

We may log data at a 1 min interval, but what to do if we want the data for whole year. For this we might simply store the data in a fixed interval but this will take so much of disk space. Also time to interpolate such large data will also be more. Using CF this problem is solved. We can define interval and Consolidation function which could be average, minimum, maximum, total etc to build our RRD. In Fig. 2 we see the significant requirement of consolidation function. We intent 1 data point every 5 min so it should be a very large database now consolidation function will keep all the data pints under the curve. This curve is known as RRDtool's data curve.
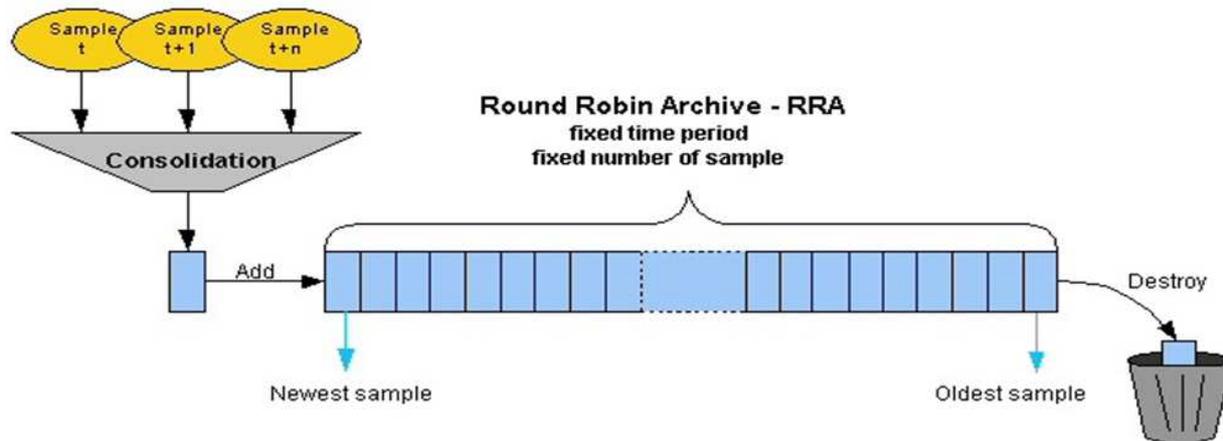
■■■

Fig. 3. RRD's Mechanism to plot Data Points

Round Robin Archives (RRA)

Round Robin Archives (RRA) stored Data values of the same consolidation function (http://oss.oetiker.ch/rrdtool/). If we like to store some 100 values in some 2 min interval, RRDTool will allocate the space accordingly. There is a header stored which stores the information about slots.

In Fig. 3 we see how archival of data is done in RRDTool database. The sampled data is consolidated using some consolidation function. New values are written to the Round Robin Archive in round robin manner. The oldest data is destroyed. This will automatically limit the history to the last 100 values. Using RRA we ensure that the RRD does not grow large and old data is eliminated. Consolidation feature will help us to keep data for a very long time. We can use different CF for different requirement like the minimum Traffic on the LAN, the maximum Temperature of Nuclear reactor, the total minutes of down time, etc. Unknown.

RRD stores data at a constant interval. Sometimes no new data is available, so with RRDtool we can handle such situations by storing an "UNKNOWN" value into the database. When consolidating a dataset, the amount of "UNKNOWN" data values is accounted for and when a new consolidated value is ready to be written to its Round Robin Archive (RRA) (Rose and McCloghrie, 1990), a validity check is performed to make sure that the percentage of unknown values in the data point is above a configurable level. If not, an "UNKNOWN" value will be written to the RRA.

Plotting and Graphing

RRDtool allows us to consolidate data as it becomes older. RRDtool stores data in a binary file. It has dump function, so the data could be dumped in an xml file. Beautiful graphs could be built using some of the following functions. RRDtool allows reports in numerical and graphical format. The graphing feature is fully configurable. Size, color and contents of the graph can be defined freely.

Implementation and Results

Historic information about the network is very important to know the trend of traffic. We can get the base how the network works in normal and abnormal conditions. We can then compare the change in network using this baseline. We want to analyze the trend of network. To estimate the future needs or performance of the network, this metric prepared by the base lining is useful. We can monitor device health and network problems on the basis of the metric. Some proactive measures are required to manage faults and monitor the network.

RRD Tool is responsible of generating graphs in layer 3. RRD Tool stores the time series data a like network bandwidth, temperatures, CPU load in a circular database (Rose and McCloghrie, 1990). The following steps are used in account of graphing using RRD Tool.

Initialize the Database

To create a database we use rrdcreate command.

The below example is made from a small project on (juniper.net) where shell script is generated to create the rrd.
Rrdcreate — step60 - NDS:pl:GAUGE:120:0:100 DS:rtt:GAUGE:120:0:10000000
RRA:MAX:0.5:1:1500
"step 60"- Then user has to specify the amount of time after which we want our data base to be updated

"start N" – this command is used to start the timer Now.

"DS: pl: GAUGE: 120: 0: 100"- DS stands for Data-Set and PL is Packet Loss. DS says that this is a data set. Pl is packet loss. GAUGE is absolute data. 120 is the timeout value in seconds. The accepted value lies between 0 and 100.

"DS: rtv: GAUGE: 120: 0: 10000000"- rtv is "round trip time value". 120 is the timeout time in seconds. The accepted value lies between 0 and 10000000.

"RRA:MAX:0.5:1:1500"- RRA is used for archival of the data points. MAX accept the maximum value if multiple values are available over multiple "steps". 0.5 is an internal resolution value, "1" specifies how many steps should be averaged before storing. One step is equals to one database value. 1500 is how many "steps" we will store in the database.

The Table 1 is created using the content on official website of RRDTool RRD tool Home Page. (http://oss.oetiker.ch/rrdtool/; Network traffic monitoring software).

Collect the Data Sets over Time

Figure 4 discusses about the stack inspired OnDisk structure of RRDTool. The topmost filed is Header, Data Structure definition followed by Round Robin Archival Definition, followed by Live Seed. Then we see the functions applied on data are stored in this tsack. We update the rrd after every specified interval. A cron job can be used to run at background and it will be used to enter the data periodically into the database. We wrote a shell script to update the database from time to time where the cron job was called. It is stored in cron.d as it is running daily.

Create the Graph

We want to create the graph for graphical representation of the collected data. We also want to run this step using a cron job. RRD Tool aims to handle time-series data like network band-width, temperatures, CPU load, etc. So we used a shell script to generate following graphs using few shell scripts. The graphs below are the graphs of memory utilization on the workstation. The graphs are daily, weekly, monthly and yearly memory utilization. These graphs can be saved in PNG format.

The user can get graphs of other information like bandwidth, CPU utilization and memory utilization of network devices. The user has to just give the IP Address and the community string of his SNMP, if it is SNMP v1 and for SNMP V2 and V3 he is supposed to give the Security name, Security level, Authentication Protocol, Authentication key and Encryption key Later the information of traffic on the interface of the network devices will be logged by the RRD Tool.
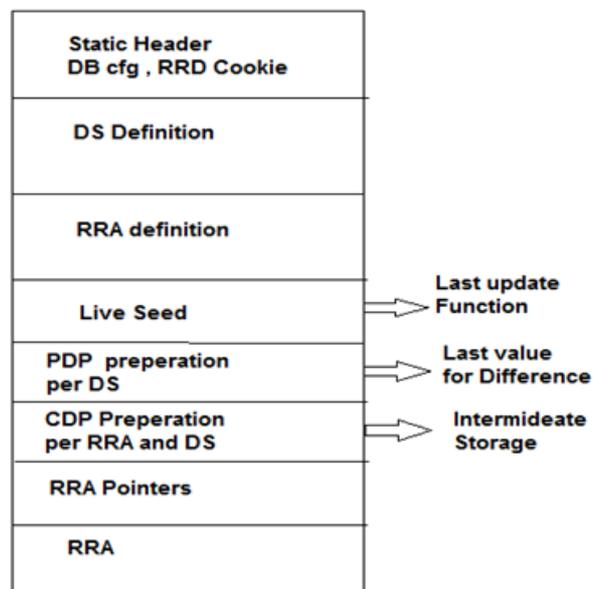

Fig. 4. On disk structure of RRDtool

Table 1. RRDTool Commands and description

1.	Rrdcreate	Setup a new Round Robin Database (Network traffic monitoring software)
2	Rrdupdate	Store a new data values into an RRD
3	Rrdupdatev	Operationally equivalent to update except for output
4	Rrdgraph	Create a graph from data stored in one or several RRDs. Apart from generating graphs, data can also be extracted to standard output
5	rrddump	Dump the contents of a RRD in plain ASCII. In connection with restore, we can use this to move an RRD from one computer architecture to another
6	Rrdrestore	Restore an RRD in XML format to binary RRD
7	Rrdfetch	Get data for a certain time period from a RRD. The graph function uses fetch to retrieve its data from an RRD

■■■

Conclusion and Future Work

We get the information of the network in a raw form when we are just querying the network devices. RRDtool can be used for storing the statistics received from the querying. Low cost tool like RRDTool can be used for Real-time visualization of network status in a graphical format. We can plot disk usage, memory utilization, spam mails etc using RRD Tool in graphical format. The data is stored in a way that system storage footprint remains constant over time. This avoids resource expensive purge jobs and reduces complexity. MySQL does yet provide this kind of storage engine. RRDTool stores the resampled version of the data rather than storing it directly.

A Round Robin Database can be used in Network monitoring. Using this tool the gap can be bridged between the network administrator and the problems lying in the network. So many monitoring systems are being implemented using RRD Tool. Open Source and well known Network Monitors like CACTI, NAGIOS, PRTG, MRTG (Zeng and Wang, 2009) are using it for graphing and storing purpose.

Acknowledgment

I want to acknowledge institute for plasma research Ahmedabad to support me with the resources. I would like to thank the head Dr. Av Ravi and all the members of computer and it department for guiding me and helping me with the project.

Author's Contributions

Sweta Dargad: First Author is responsible in research work and development of manuscript.

Manmohan Singh: Co-author is responsible for publication and has been a resourse provider

Ethics

There are no ethical issues in publishing and giving open Access.

References

Harrington, D., R. Presuhn and B. Wijnen, 2002. An architecture for describing Simple Network Management Protocol (SNMP) management frameworks. RFC.

http://oss.oetiker.ch/rrdtool/

Network Monitoring/Management. http://www.cotse.com/tools/netman.html

Networktrafficmonitoring software. http://www.topology.org/comms/netmon.html

Rose, M. and K. McCloghrie, 1990. Structure and identification of management information for TCP/IP-based internets. RFC 1155.

RRD tool Home Page. http://oss.oetiker.ch/rrdtool/

Zeng, W. and Y. Wang, 2009. Design and implementation of server monitoring system based on SNMP. proceedings of the International Joint Conference on Artificial Intelligence, Apr. 25-26, IEEE Xplore Press, pp: 680-682. DOI: 10.1109/JCAI.2009.34

■■■