
A Monitoring System for the BaBar INFN Computing Cluster



Moreno Marzolla

*Università "Ca' Foscari" di Venezia
and
INFN, Padova*

Valerio Melloni

*Dip. Matematica,
Università di Ferrara*

marzolla@pd.infn.it



Presented by:

Fulvio Galeazzi
INFN, Padova

fulvio.galeazzi@pd.infn.it

Talk Outline

- Introduction
- Motivation: Monitoring the *BaBar* Computing Farm
- *PerfMC*: A prototype of an SNMP-Based monitoring application
- Conclusions



Fulvio Galeazzi, CHEP 2003, Mar 24–28 2003

Monitoring

- A **monitor** is a tool used to observe the activities on a system
 - Collects performance measures
 - (Possibly) Analyzes the data
 - Displays the results.
- Why?
 - Measure **resource utilization** to find performance bottlenecks;
 - Characterize the **Workload**;
 - Find **model parameters**, validate models, or develop inputs for a model.

BaBar Farm @ INFN Padova



- ≈ 170 2xPIII 1.26GHz Machines, 1GB Ram, RH Linux 7.2
 - 130 Clients
 - 40 Servers
- Tape Library with a capacity of ~ 70 TB not compressed;
- Network switches, UPSes, Environmental conditioning systems,

...

Monitoring Requirements

- **Hardware Status**
 - Machine Crashes, CPU utilization, Disk I/O, Network I/O...
- **Processes status**
- **Environmental conditions**
 - Humidity, Temperature, UPS status...
- Does not need to be a real-time monitor
- The monitoring system should also be:
 - *Reasonably Scalable*
 - Efficient (low resources requirement)
 - *Flexible and customizable*
 - *Easy to configure*
 - Able to operate in background

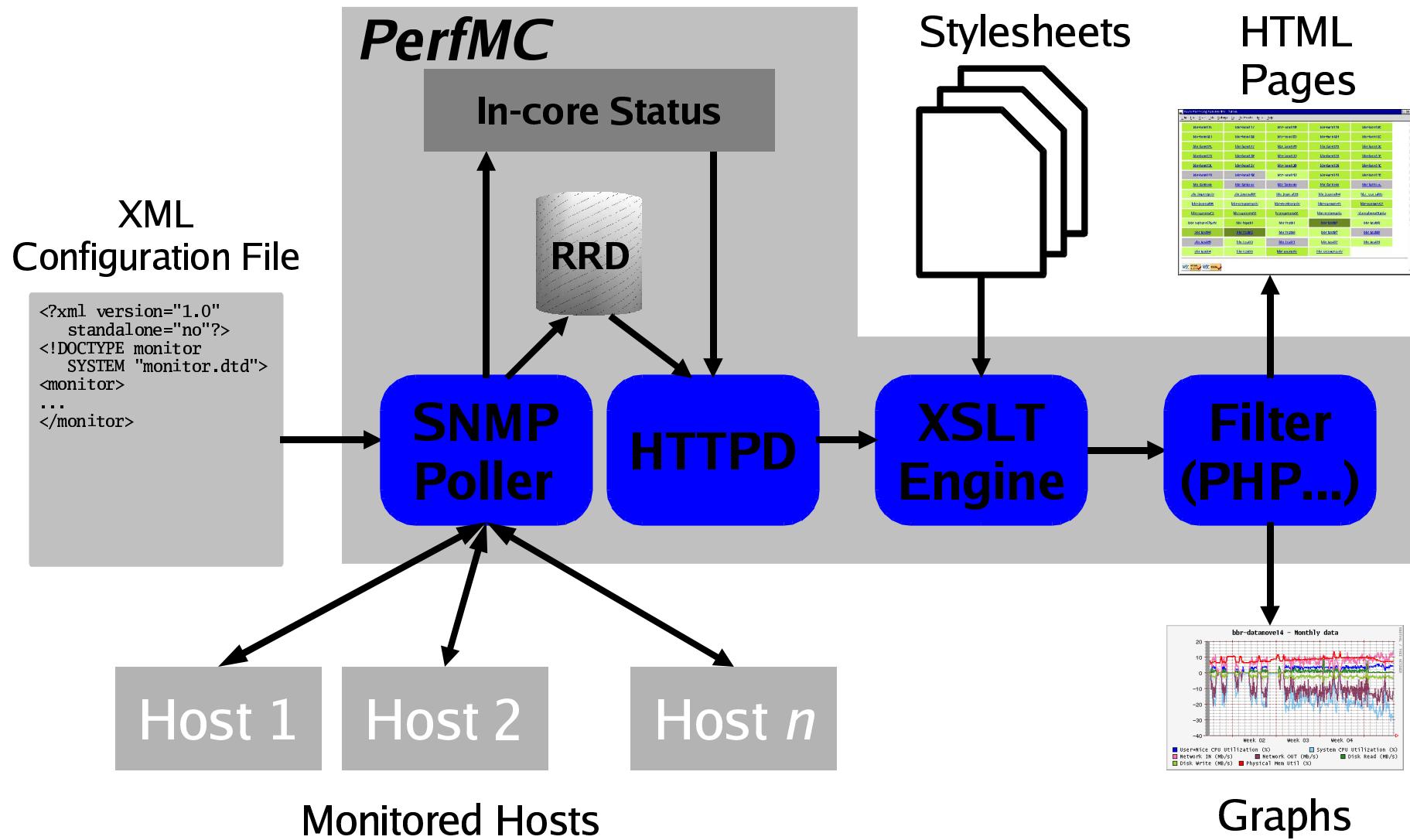
Some problems with existing tools

- Limited scalability
- Require their own daemons running on the monitored hosts
 - Can't install a daemon on a network switch, or on a tape library
- Hard to configure
- Poorly implemented
 - Heavy use of scripting languages, mixed C/Perl/shell pieces

PerfMC: a Performance Monitor for Clusters

- Characteristics:
 - Written in C
 - Asynchronous (nonblocking) parallelized SNMP Polling
 - Uses SNMPv2 Bulk Get requests
 - XML-based configuration file
 - The **RRDTool** package is used to store data and produce graphs
 - Old data have lower resolution than recent ones
 - Round Robin Databases have known, fixed size
 - Graphing capabilities are provided by the library
 - Dynamic generation of HTML pages using XSLT stylesheets and a filter (PHP...)
 - PerfMC has an embedded HTTP server

PerfMC Architecture



Sample HTML Output

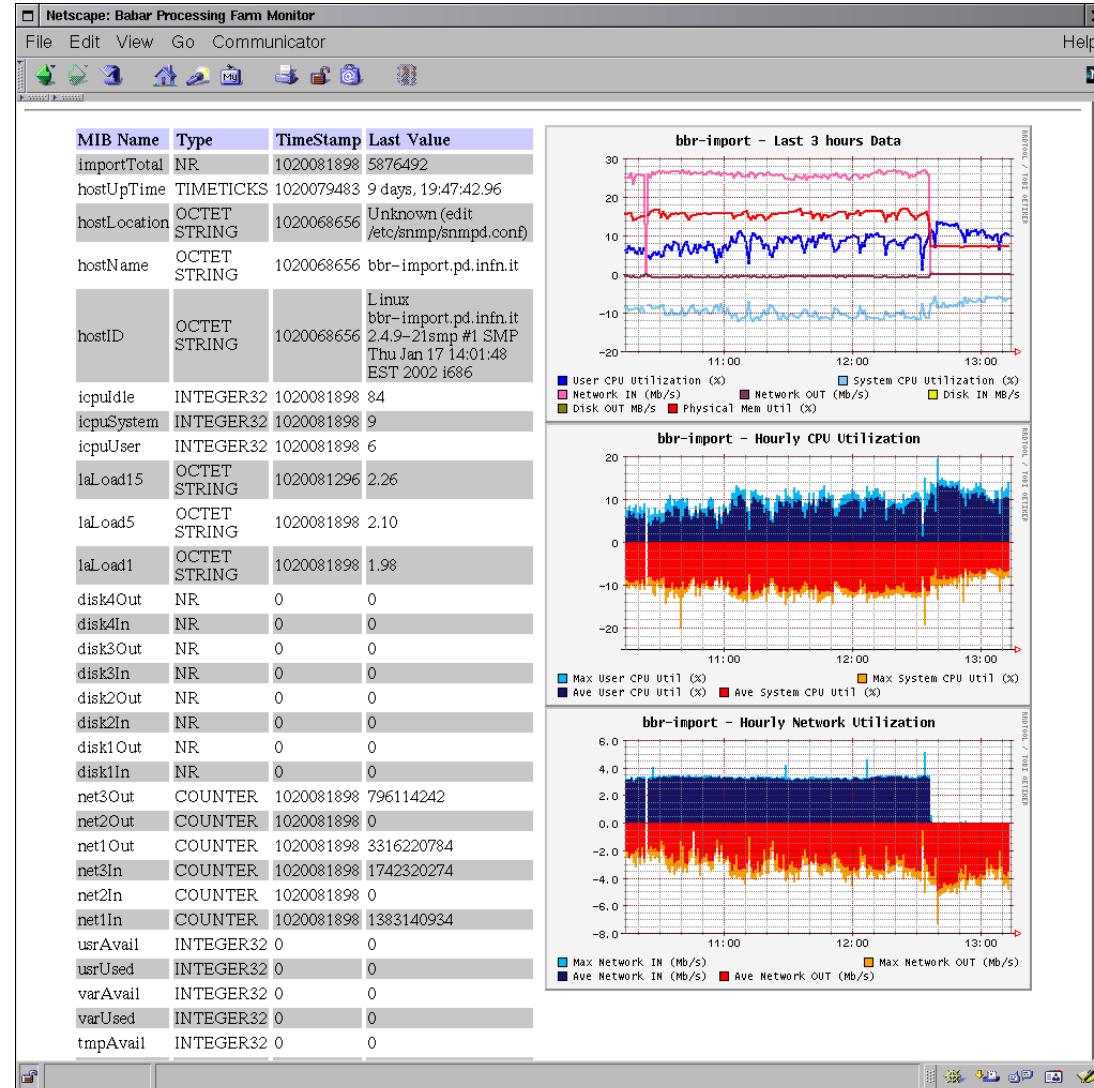
Netscape: Babar Processing Farm Monitor

File Edit View Go Communicator Help

Bookmarks Location: Moreno Home BaBar Home Linux.ORG Goog

Farm Overview

SNMP Status	Machine
NR	bbr-cndserv01
OK	bbr-datamove01
OK	bbr-farm001
NR	bbr-farm002
NR	bbr-farm003
NR	bbr-farm004
NR	bbr-farm005
NR	bbr-farm006
NR	bbr-farm007
NR	bbr-farm008
NR	bbr-farm009
OK	bbr-farm010
OK	bbr-farm011
NR	bbr-farm013
NR	bbr-farm014
NR	bbr-farm015
OK	bbr-farm016
NR	bbr-farm020
OK	bbr-import
OK	bbr-tape01
NR	bbr-test01
OK	bbr-user
OK	catbb1

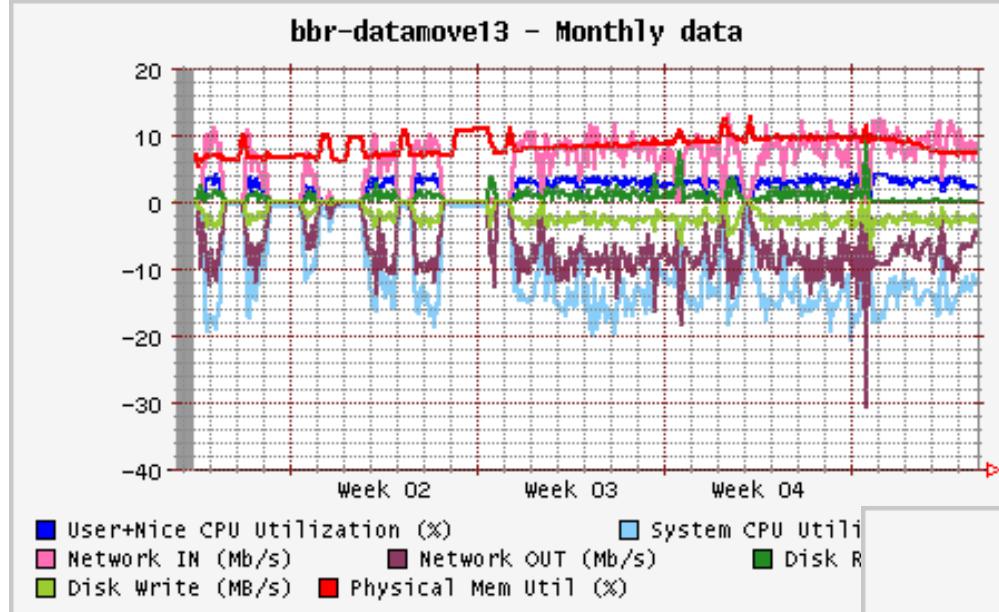


Another example

Babar Processing Farm Monitor - Galeon				
File	Edit	View	Tab	Settings
Go	Bookmarks	Tools	Help	
bbr-farm116	bbr-farm117	bbr-farm118	bbr-farm119	bbr-farm120
bbr-farm121	bbr-farm122	bbr-farm123	bbr-farm124	bbr-farm125
bbr-farm126	bbr-farm127	bbr-farm128	bbr-farm129	bbr-farm130
bbr-farm131	bbr-farm132	bbr-farm133	bbr-farm134	bbr-farm135
bbr-farm136	bbr-farm137	bbr-farm138	bbr-farm139	bbr-farm140
bbr-farm141	bbr-farm142	bbr-farm143	bbr-farm144	bbr-farm145
bbr-farm146	bbr-farm147	bbr-farm148	bbr-farm149	bbr-farm150
bbr-importpriv	bbr-journal02	bbr-journal03	bbr-journal04	bbr-journal05
bbr-journal06	bbr-mngservpriv	bbr-monitorpriv	bbr-oprserv01	bbr-oprserv02
bbr-oprserv03	bbr-oprserv04	bbr-oprserv05	bbr-restorepriv	bbr-sqlserv01priv
bbr-sqlserv02priv	bbr-tape01	bbr-test01	bbr-test02	bbr-test03
bbr-test04	bbr-test05	bbr-test06	bbr-test07	bbr-test08
bbr-test09	bbr-test10	bbr-test11	bbr-test12	bbr-test13
bbr-test14	bbr-test15	bbr-userpriv	bbr-webservpriv	

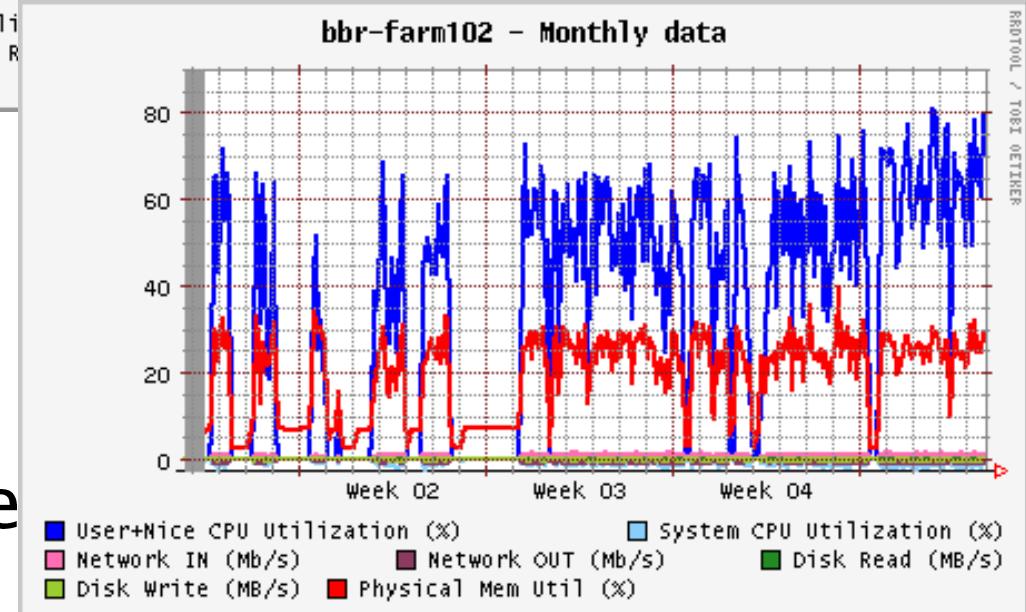
 

Some Plots

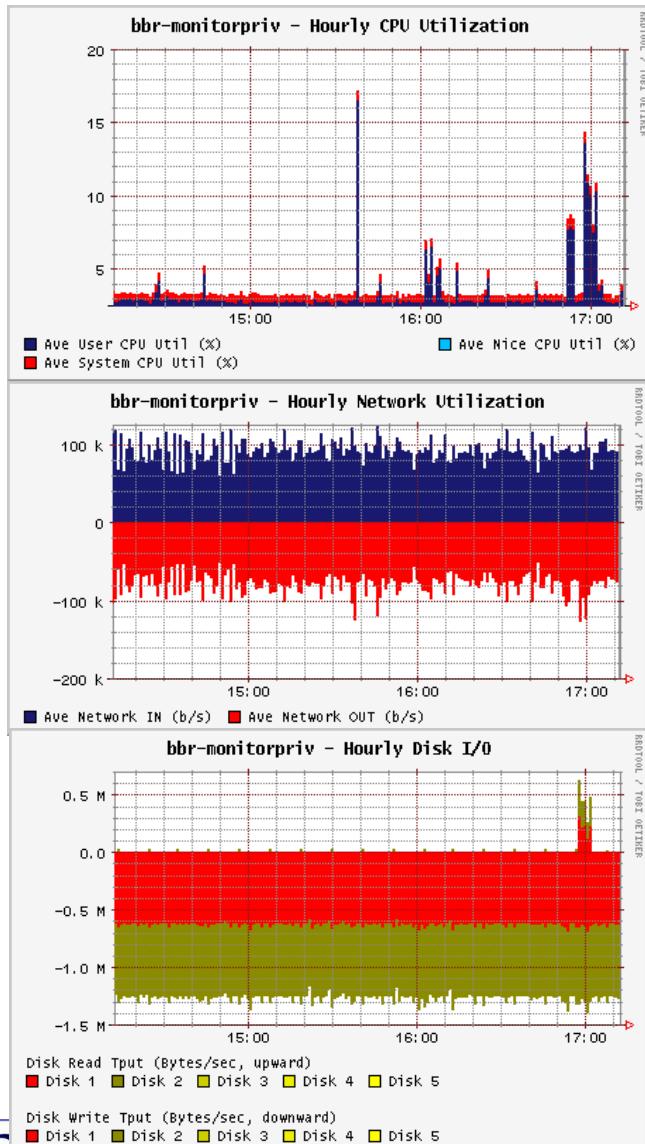


Database
Server

Client
Machine



PerfMC Performances



Reasonably low CPU Utilization (< 2%)

Reasonably low Network Utilization (11 KB/s)

Not-so low Disk Utilization (1.2 MB/s)

Conclusions

- Monitoring a large computing cluster is a highly nontrivial task.
- Many available monitoring tools exist, but many of them are not adequate for large distributed systems.
- We are trying to build a general-purpose SNMP and XML-based monitoring tool.
- A prototype exists and is working well

Future work

- Alarms are not currently implemented, but are at the top position of the to-do list
- At the moment there are no serious problems.
 - PerfMC is running on the production cluster (≈ 170 machines)
 - Clearly cannot scale forever
 - Single point of failure
- No attention has been put on *security*
 - Could easily be extended for SNMPv3
 - Not a priority. Our cluster is on a private network

Bibliography

- Moreno Marzolla, *A Performance Monitoring System for Large Computing Clusters*, proceedings of PDP2003, Genova, Italy, Feb 5–7, 2003
- W. Stallings, *SNMP, SNMPv2, SNMPv3 and RMON 1 and 2*, third edition, Addison-Wesley, 1999
- Grid Performance Working Group
<http://www-didc.lbl.gov/GridPerf/>
- RRD Tools Home Page
<http://people.ee.ethz.ch/~oetiker/webtools/rrdtool>
- BaBar Farm home page (will contain PerfMC)
<http://bbr-webserv.pd.infn.it:5211/farm/index.html>
- BaBar Farm Monitoring Page
<http://bbr-monitor.pd.infn.it:5211/monitor/html/index.html>

Backup Slides

Example of XML Configuration File

```
<?xml version="1.0" standalone="no"?>
<!DOCTYPE monitor SYSTEM "monitor.dtd">

<monitor numconnections="50" pmclogfile="/monitor/pmc.log"
          httpdlogfile="/dev/null" rrddir="/monitor"
          htmldir="/monitor/html" pmcverbosity="3" >

  <host name="localhost" tag="client">
    <description>This is a sample client machine</description>
    <miblist>
      <!-- list of mibs to monitor -->
    </miblist>
    <archives>
      <!-- RRD layout -->
    </archives>
    <graphs>
      <!-- Graph definitions here -->
    </graphs>
  </host>

</monitor>
```

Sample XML configuration file (MIB)

```
<miblist>
  <mib id='tempMB' name='1.3.6.1.4.1.2021.13.16.2.1.3.1'/>
  <mib id='tempCpu1' name='1.3.6.1.4.1.2021.13.16.2.1.3.2'/>

  <!-- .iso.org.dod.internet.private.enterprises.ucdavis.systemStats.ssCpuRawUser.0 -->
  <mib id='cpuUser' name='1.3.6.1.4.1.2021.11.50.0' type='COUNTER'/>

  <!-- .iso.org.dod.internet.private.enterprises.ucdavis.systemStats.ssCpuRawSystem.0 -->
  <mib id='cpuSystem' name='1.3.6.1.4.1.2021.11.52.0' type='COUNTER'/>

  <!-- .iso.org.dod.internet.private.enterprises.ucdavis.systemStats.ssCpuRawNice.0 -->
  <mib id='cpuNice' name='1.3.6.1.4.1.2021.11.51.0' type='COUNTER'/>

  <!-- .iso.org.dod.internet.mgmt.mib-2.interfaces.ifTable.ifEntry.ifInOctets.2 -->
  <mib id='net1In' name='1.3.6.1.2.1.2.2.1.10.2' type='COUNTER'/>

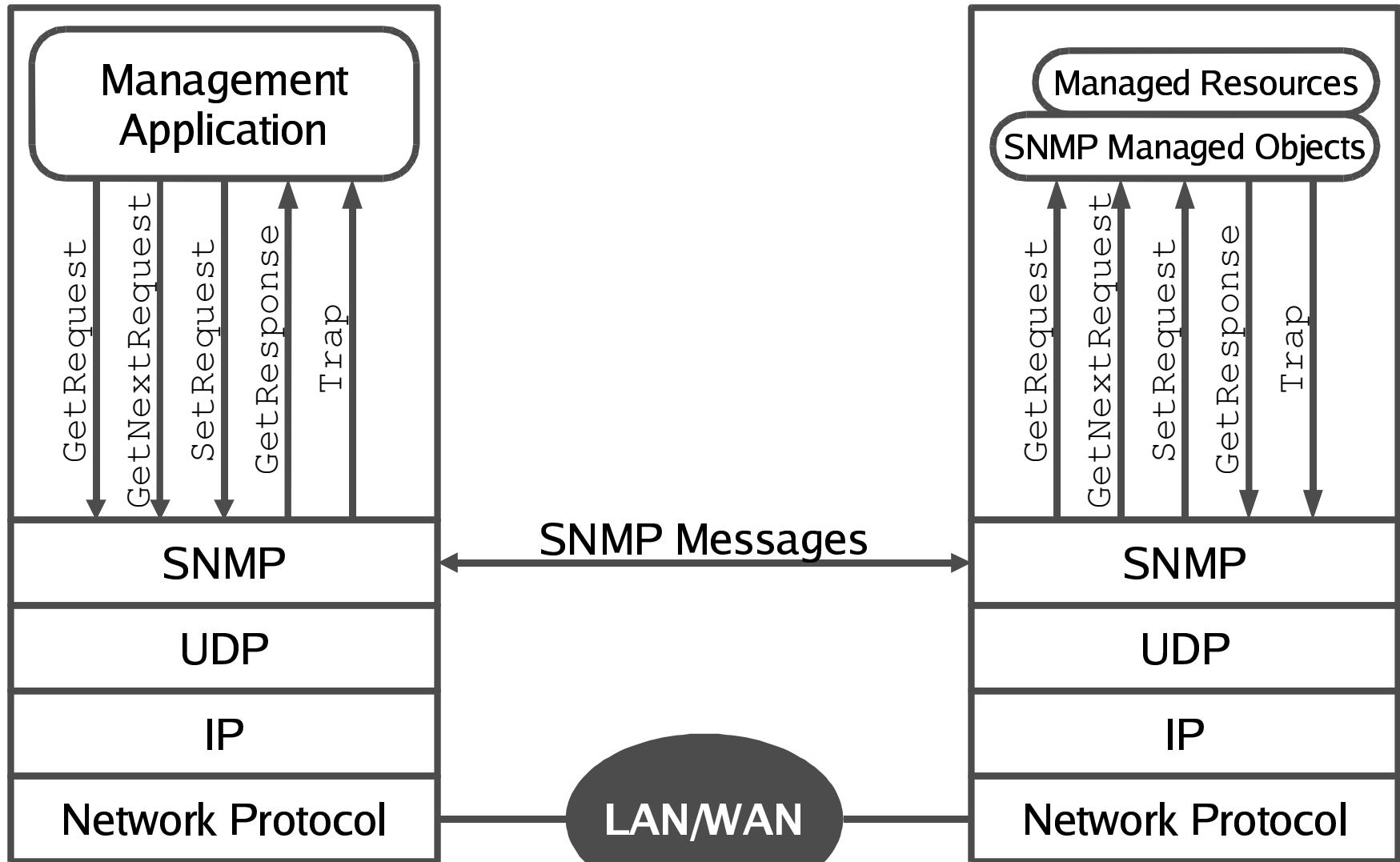
  <!-- .iso.org.dod.internet.mgmt.mib-2.interfaces.ifTable.ifEntry.ifOutOctets.2 -->
  <mib id='net1Out' name='1.3.6.1.2.1.2.2.1.16.2' type='COUNTER'/>

</miblist>
```

Example of XML Status Dump

```
<?xml version="1.0"?>
<hosts>
    <host name="localhost" status="NR">
        <mibs>
            <mib id="availSwap" lastUpdated="1018016033">1052248.000000</mib>
            <mib id="totalSwap" lastUpdated="1018016033">1052248.000000</mib>
            <mib id="totalMem" lastUpdated="1018016033">917080.000000</mib>
            <mib id="cachedMem" lastUpdated="1018016033">7128.000000</mib>
            <mib id="bufferMem" lastUpdated="1018016033">35052.000000</mib>
            <mib id="sharedMem" lastUpdated="1018016033">0.000000</mib>
            <mib id="freeMem" lastUpdated="1018016033">833800.000000</mib>
            <mib id="cpuSystem" lastUpdated="1018016033">137587.000000</mib>
            <mib id="cpuUser" lastUpdated="1018016033">13581.000000</mib>
            <mib id="tempCpu2" lastUpdated="1018016033">24500.000000</mib>
            <mib id="tempCpu1" lastUpdated="1018016033">25000.000000</mib>
            <mib id="tempMB" lastUpdated="1018016033">33000.000000</mib>
        </mibs>
        <graphs>
            <graph id="hourly.png" title="Hourly data"/>
        </graphs>
    </host>
</hosts>
```

More on SNMP



Round Robin Databases

