



Performance Training SUN Users

Dimitri KRAVTCHUK
dimitri@sun.com
Benchmark Team
Paris Sun Solution Center



Before we start...

- Few words about SSC :-)
- Paris <=== 10Mbit, 20ms latency ===> LLG



SSC Locations

- North America
 - > USA: Hillsboro, Broomfield, McLean, Menlo Park
- Latin America
 - > Sao Paulo, Brazil; Ft. Lauderdale, Florida; Mexico City, Mexico
- Europe
 - > Edinburgh, Frankfurt, Madrid, Manchester, Milan, Munich, Paris, Walldorf
- Asia
 - > Bangalore, India; Beijing, China; Hong Kong; Seoul, Korea; Singapore; Taipei, Taiwan; Tokyo, Japan;
- Pacific
 - > Sydney, Australia



Sun Solution Center Is Near You



ASSCs in BLUE
SSCs in BLACK

United States

San Francisco Bay Area, CA
Hillsboro, OR
Broomfield, CO
Mc Lean, VA
Chicago, IL – Diamond Management
Plano, TX - EDS
College Park, MD - Univ of Maryland
Pittsburgh, PA – Deloitte Consulting

Latin America

Ft. Lauderdale, FL, USA
Mexico City, Mexico
Sao Paulo, Brazil

Europe / Middle East / Africa

Edinburgh, Scotland, UK
Manchester, UK
Warrington, UK - Avnet
Paris, France
Frankfurt, Germany
Munich, Germany
Walldorf, Germany
Milan, Italy
Madrid, Spain
Göteborg, Sweden – Inserve Technology
Helsinki, Finland – ArrowECS
Tallin, Estonia - Microlink
UAE - Tech Access

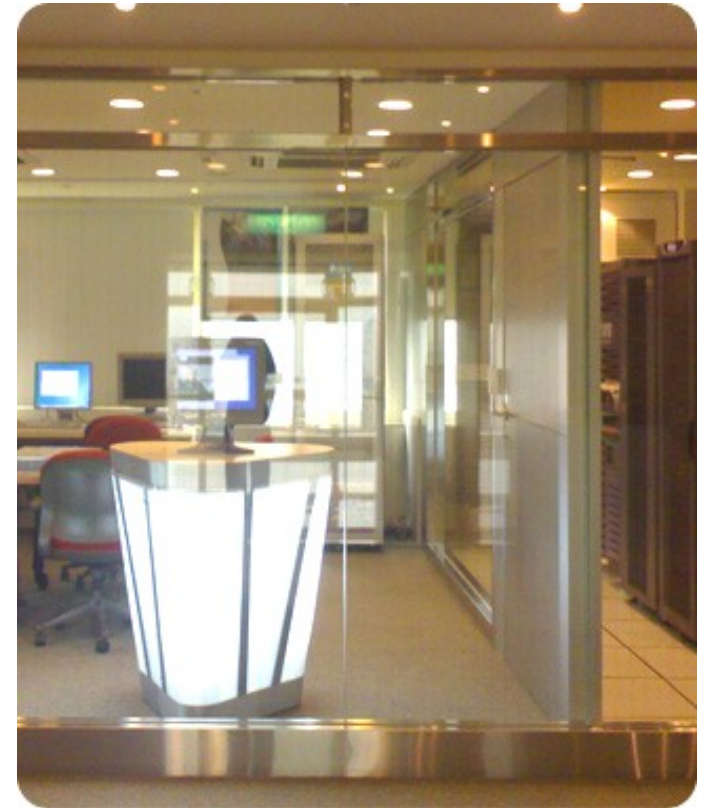
Asia Pacific

Bangalore, India
Bangalore, India - Wipro
Beijing, China
Hong Kong, China
Shenyang, China - Neusoft
Seoul, Korea
Singapore
Singapore - Ingram Micro
Sydney, Australia
Sydney, Australia – Express Data
Tokyo, Japan

Sun Solution Center

Benchmark and Performance Characterization

- Architecture design
- High-end performance and scalability
(servers, storage)
- Performance characterization
- Competitive benchmarks
- Internal product BU benchmarks
- Performance tuning
- Customer/Partner benchmarks
- Customer briefings



Sun Solution Center

Partner Solution Center

Architecture design and validation

Portfolio management and solutions offerings

Customer/Partner Proof-of-Concepts

End-to-end software development for live customers

Industry solutions development and showcase

Building of horizontal/biz solutions (eg: IdM, Security ... etc.)

Business innovation and compliance (SOX, HIPA ... etc.)

Demos, solution showcase

To know more

<http://www.sun.com/solutioncenters>

Test for success.

We assembled the best team in the industry to assess unique business solutions.



Overview

Services

Locations

Get Started

At a Glance | Welcome Letter | FAQs



"Most of our customers share two characteristics; they believe in the power of the community to solve challenging problems, and they believe that technology is a competitive differentiator for their business. The Sun Solution Centers bring together state-of-the-art technology and expertise in simulated environments where our clients can envision, build, and test innovative business solutions." Jonathan Schwartz, President and CEO.

What can Sun Solution Centers do for you?



The goal of the Centers is to minimize your risk, justify your expense, and shorten time to deployment of your new business solutions by providing the tools you need to 'test before you invest'. We do this by offering Sun and Sun partner access to in-depth expertise in technologies, industries, and applications in collaborative, state-of-the-art

Working with Sun Solution Centers

» How to Get Started

Considering a new business solution? Interested in exploring in-depth what the power of Sun can do for you? Get started by contacting your Sun Account Manager or Systems Engineer. They can initiate the process by discussing your needs with you and then requesting an engagement with the Sun Solution Center.

This Month's Top 5 Requested Services

- Finance Industry POC
- Telco POC
- SAP Sizing
- Customer Workshop
- HPC Performance Consulting

» See all Services



Authorized Sun Solution Centers

Find out where they are.

The Goal of this Training...

- IS NOT...

- > is not to teach you live...
- > is not a hypnotic séance in group...
- > is not a “religion” question...
- > is not a UserGroup recruit...
- > etc. etc. etc.

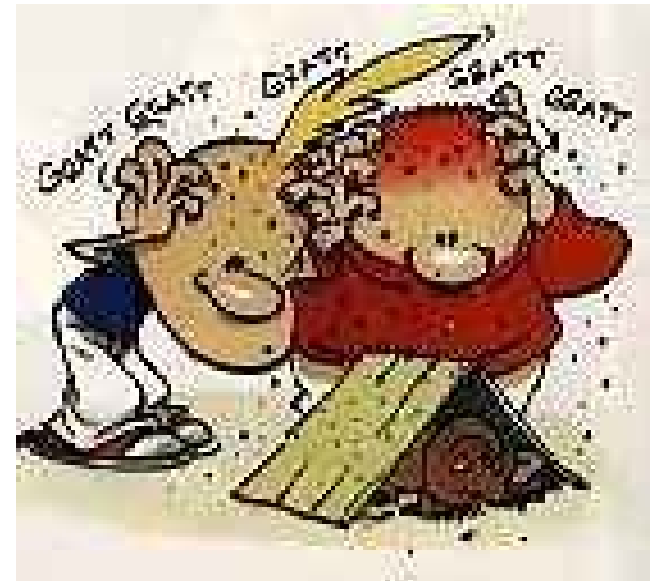


Keep in mind...

- 1. NOBODY KNOWS EVERYTHING!
- 2. NOBODY KNOWS EVERYTHING!
- 3. NOBODY KNOWS EVERYTHING!
- 4. NOBODY KNOWS EVERYTHING!
- ...
- GOTO #1
- And probably what I'm saying right now is already obsolete... :-)
- Stay tuned & thinking! :-)

The Goal of this TOI...

- IS TO SHARE!!!!
- INVITE YOU TO DISCUSS!
- ASK QUESTIONS!
- HAVE FUN! :-)



Agenda

- Part1: Performance problems & analyzing
 - > Network... + Q & A
 - > I/O... + Q & A
 - > Processes, CPU, RAM, etc.... + Q & A
 - > UFS, VxFS, QFS, ZFS, Dtrace, CMT, etc...
 - > Q & A
- Part2: All about dim_STAT :-)
 - > Idea, Architecture Overview, Getting Started...
 - > Analyzing, Reports, By Example Demo
 - > Live Demo ?
- Q & A



Part1: Preface...



Preface: Close Look...



Preface: Close Look zoom--



Preface: Close Look zoom-- x10



Sun Platforms...

Intel / AMD 64



SPARC64-VI/VII



**UltraSPARC T2/T2+
“CoolThreads”**



Part1: Performance...

QUESTION: *Do you THINK we have the best products in the world?...*

Part1: Performance...

QUESTION: *Do you THINK we have the best products in the world?...*

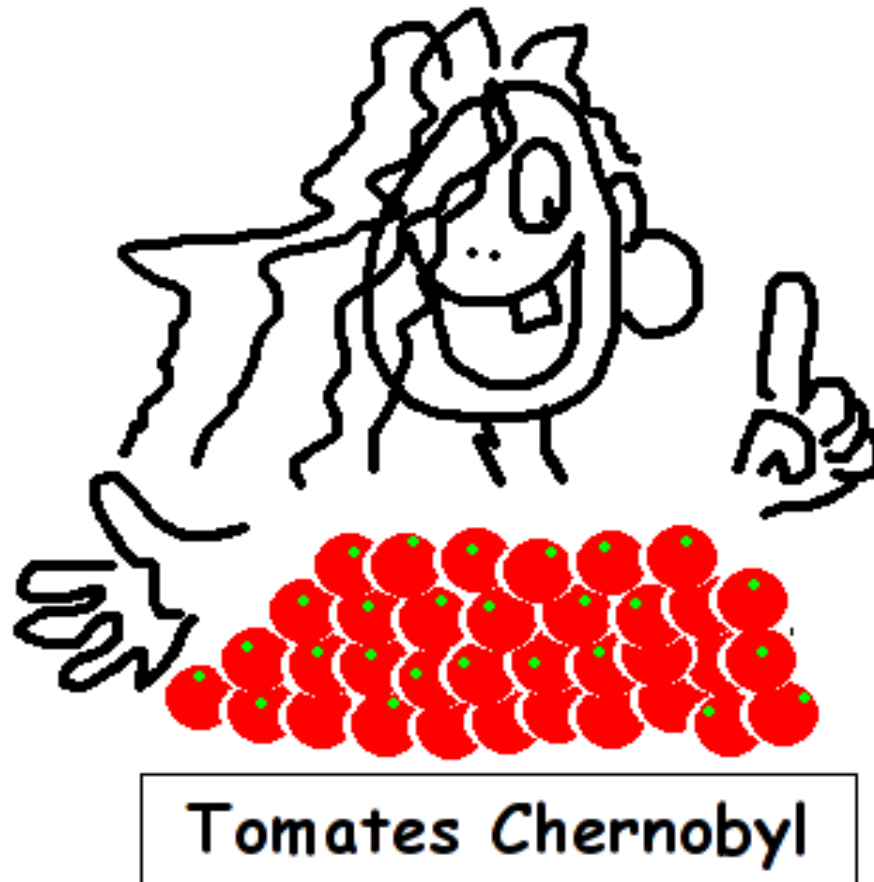
Dimitri: *Most of our products have a huge potential for improvement!
And they are often the best because our competitors are doing even
WORSE!...*



Performance Analyzing...

- Break your mind...
- Reconsider all pieces of your platform...
- Go to the ROOT of the problem!
- Understanding of the problem = 50% of the solution!

Break your mind...



(clim)

Performance: from worse to better...

- **Application :-)**
- Network
- I/O Subsystem & Storage
- Processes
- CPU
- Memory
- Other...
- NOTE: “health altitude”...



Performance: from worse to better...

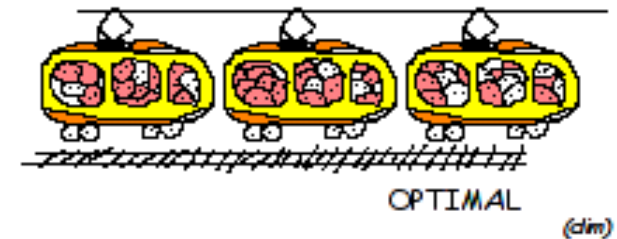
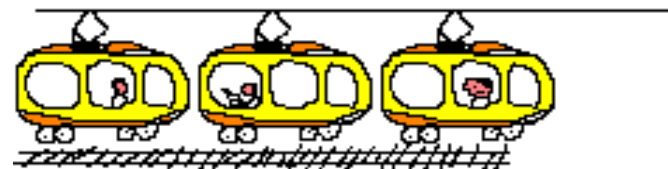
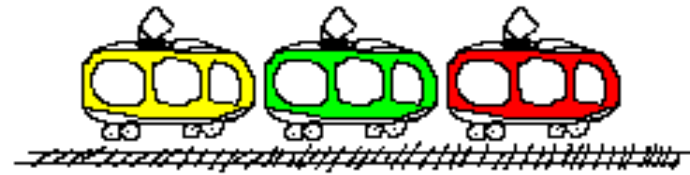
- Application :-)
- ==> Network



Network

- Network is the computer
- STREAMS...
- Fire Engine
- Improved all the time!

Network Internals...



Network Monitoring

- kstat – main source
 - > Before Solaris 10 freezing traffic(!) => upgrade!!!
- tools: netstat, nicstat, netLOAD
- netLOAD
 - > packets/sec
 - > bytes/sec
 - > errors in/out
 - > collisions
 - > nocanput (!)
- Known limits
 - > 100Mbit => 12M/s, Gbit/SPARC => 50-60K pack/s, etc...

Network tuning...

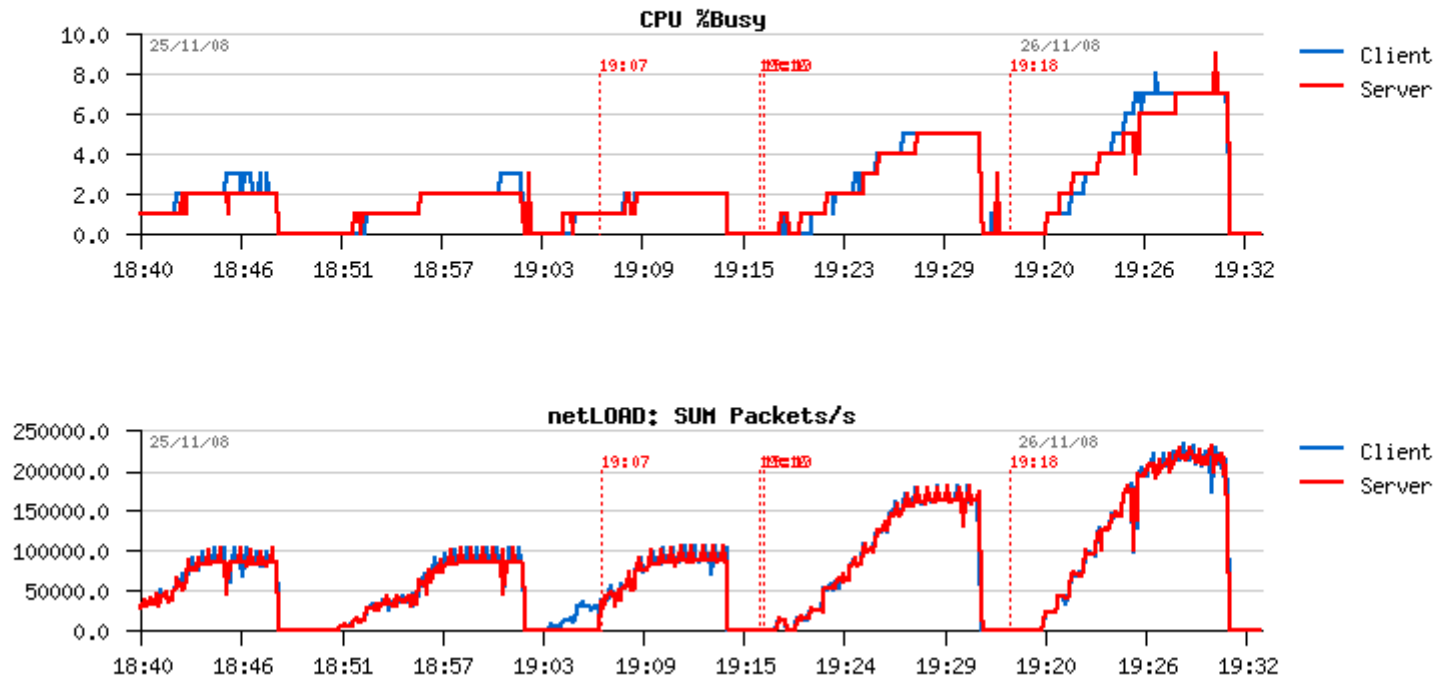
- Goal => ping-pong packets/s vs throughput?
 - > interrupts vs queued mode (set ce:ce_taskq_disable=1)
 - > mpstat=> intr => psradm -i + psrset
 - > new NICs are CPU-aware
- Queue size
 - > ndd -get /dev/tcp tcp_conn_req_max_q 1024 (def:128)
 - > ORACLE: listener queue!
- Deferred ACK / grouping/ NO_DELAY
 - > ndd -set /dev/tcp tcp_deferred_ack_interval 10 (def:100)
- Jumbo frames
- IP @Sol10 today's limit: **CPU power**

Packets/s with 1x, 2x & 3x NIC

Database: [M9K_dbSTRESS_Dimitri]

[\[Home \]](#) [\[Preferences \]](#) [\[LOG Admin \]](#)

dim_STAT Multi-Host Analyzer



Important sources

- solarisinternals.com ==> Networks
 - > Tons of tuning, platform-specific, etc.
- SunSolve :-)

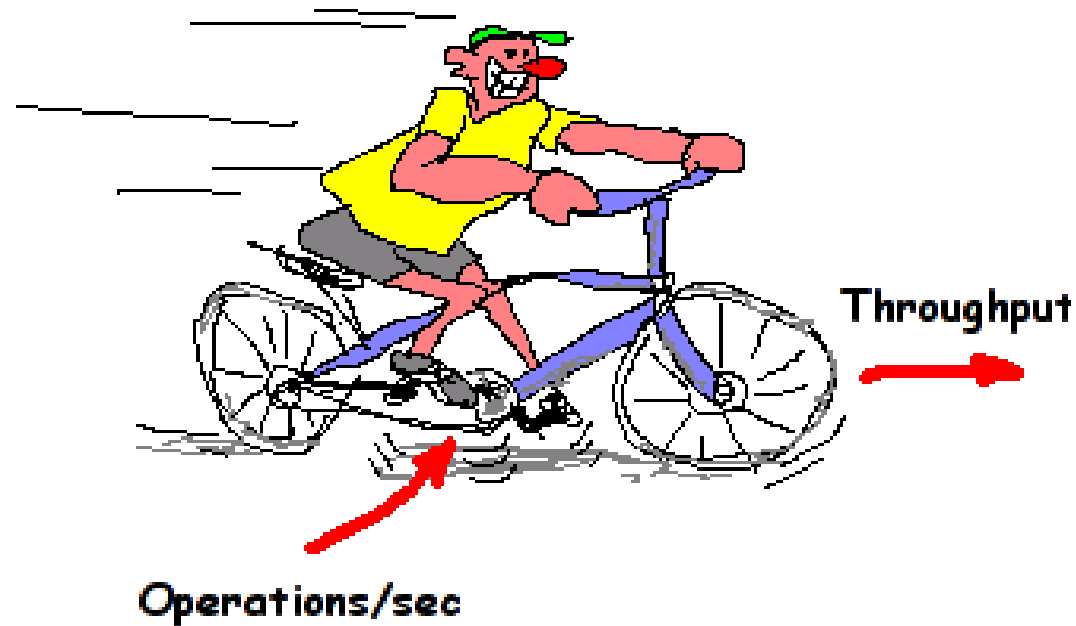
Performance: from worse to better...

- Application :-)
- Network
- ==> I/O Subsystem & Storage



I/O Performance Overview

I/O Internals...



(dim)

I/O Components and Levels

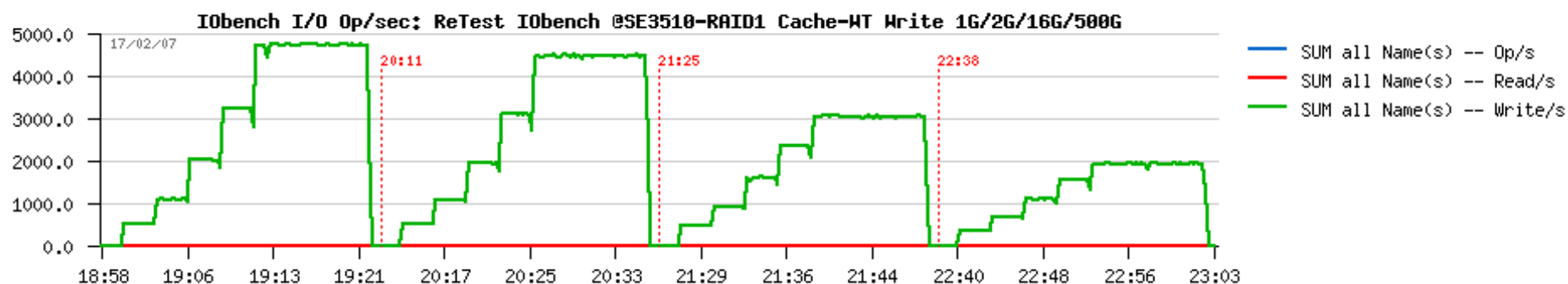
- System
 - > UFS, VxFS, QFS, ZFS, ...
 - > SDS, VxVM, ZFS, ...
 - > MPXIO, VxDMP, ...
- Storage
 - > Sun / HDS / EMC
 - > Huge Box vs Nx Small Boxes...
 - > etc...

I/O Analyze

- Type
 - > Seq.Read: prefetch
 - > Seq.Write: deferred cache flush
 - > Rnd.Write: storage cache
 - > Rnd.Read: WORSE CASE!
 - > Seek Time is IMPORTANT!
- Main info: iostat -xn
 - > 100% busy ==> Really?...
 - > actv, wait, srv.time, wait.time
 - > wait I/O => blah-blah-blah...
- DtraceToolkit: iopattern

Seek Time...

- Random I/O...
- NOTE: what about huge Oracle datafiles?..



I/O Reproduce

- Test kits
 - > VDBench, FileBench, IObench, etc.
- Test Case
 - > Check Storage! => RAW device (MPXIO off, DMP off)
 - > Workload simulation
 - > Background load?
- MPXIO, VxDMP
 - > SunSolve! :-)
 - > Ex: Sol10 U4, MPXIO-off: 5.000ops => 20.000ops

Silly Question:

- My application is doing random I/O
- With random Write-Only operations my storage box reaching 30.000 writes/sec MAX
- QUESTION: what kind of degradation in writes/sec I may expect if 10% of my I/O will become random reads?...

UFS Performance

- File system creation + tuneefs
- Mount options
 - > remount
 - > logging (!) ==> MUST
 - > noatime ==> AVOID!
 - > forcedirectio
- forcedirectio
 - > removes double buffering
 - > resolves single (POSIX) lock issue
 - > benefit for concurrent writes
 - > buffered vs direct => analyze your workload!

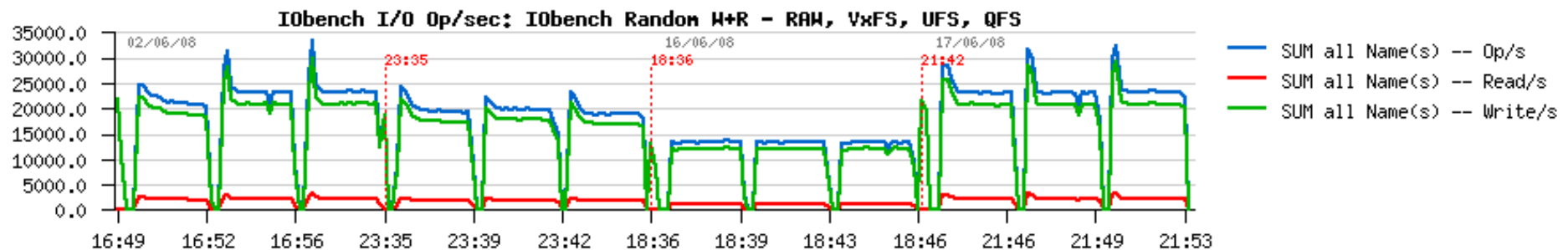
VxFS Performance

- Powered design: extents
- vxtunefs
 - > discovered direct (!) (ex. 30% gain Oracle)
 - > read/write streams number
- Direct mode
 - > mincache=direct,convosync=direct
- Quick I/O
 - > cached RAW devices
 - > Any real gain?...

QFS Performance

- Most advanced and powerful FS (on its time)
- Combines Direct WRITE + Buffered READ
 - > 50% gain over VxFS (real benchmark)
- More wide features since @Sun...
- Outstanding product!!!

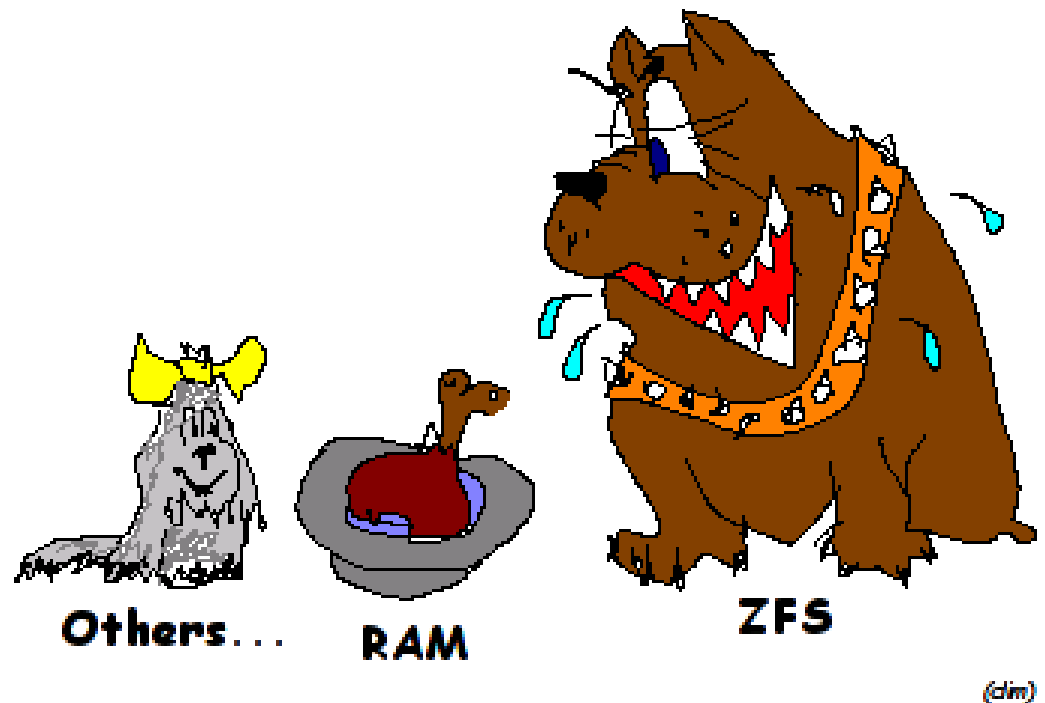
Near to RAW performance?...



ZFS Performance

- Memory hungry
 > limit it from the beginning!!!
- Read Priority (per pool)
- Block size
- Copy-on-Write
- Garbage collection
- Free space
- ARC stats
- New! write_throttle!

ZFS Internals



(dim)

SSD == Breaking rules...

- Latency: 1-2ms
- Throughput: ~20.000 Op/sec R/W (or more)
- Where to use?..
 - > **Note:** write latency on BBU storage is already 0.1ms(!)
 - > Random Read is the most benefit!
 - > ZFS: part of cache
 - > Etc. etc.etc. :-)



Performance: from worse to better...

- Application :-)
- Network
- I/O Subsystem & Storage
- **==> Processes**



Processes Activity Monitoring

- `# mv /usr/bin/top /usr/bin/top.do_not_use`
- `ps -ef :-)`
- `psSTAT -o ... -T interval`
- `prstat` (since Solaris 8)
 - > `threads (-L)`
 - > `lock time, run queue latency (-m)`
 - > `sys, usr% CPU usage (-m)`, **Sybase** => `poll()` => high `sys%`!
- Monitoring system calls
 - > `truss [-c] -p pid, DTrace (preferable)`
- Monitoring I/O
 - > `psSTAT, DTrace`

Processes Locking Monitoring

- `mpstat`
 - > spin mutex
 - > (and again, if application is well designed :-))
- `prstat -m`
 - > LCK value
- `plockstat -A -e 20 -p pid`
 - > hottest callers and objects
 - > `malloc()` is the most common case
 - > C++ multi-threaded application => 90% chance to lock
 - > fix: relink or `LD_PRELOAD` with `libmtmalloc.so`, `libumem.so`

Process related commands

- cputrack - per-processor hw counters
- pargs – process arguments
- pflags – process flags
- pcred – process credentials
- pldd – process's library dependencies
- psig – process signal disposition
- pstack – process stack dump
- pmap – process memory map
- pfiles – open files and names
- prstat – process statistics
- ptree – process tree
- ptime – process microstate times
- pwdx – process working directory
- pgrep – grep for processes
- pkill – kill processes list
- pstop – stop processes
- prun – start processes
- prctl – view/set process
- pwait – wait for process
- preap – reap a zombie process
- ...

Performance: from worse to better...

- Application :-)
- Network
- I/O Subsystem & Storage
- Processes
- **==> CPU**



Scalability...

- Scalable vs Non-Scalable Application
- Will your application run 4 times faster if -
 - > you upgrade CPU with 4 times higher frequency
 - > or increase 4 times number of the same CPUs?
- If your application is not scaling, what about your possibilities to increase performance?...
- “Free lunches are finished...”

What is doing CPU most of the time?...

- Waiting... :-)
- “In every joke there is always a part of truth”...
- Old but useful story with E10000:
 - > Customer Benchmark won with E10K 32CPU (250Mhz)
 - > Sun sold to customer E10K 24CPU but with **400Mhz**
 - > 8000Mhz => 9600 Mhz
 - > Result: Global performance slow down...
 - > Reason: runs faster with more “hands”
 - > Solution: re-tune configuration again...



CPU Monitoring

- Metrics...
 - > Sol251 vs Sol26 vs Sol7 vs Sol8 vs Sol9 vs Sol10 vs ...
- vmstat, mpstat
 - > usr, sys, idle, wait i/o => pipo
 - > run queue, blocked processes
 - > cross-calls (xcal)
- Empty cycles
 - > har
 - > cpustat, cputrack, corestat
 - > cc -fast ...
 - > strings bin/prog | grep -i workshop

Killing example

```
long long s, gethrtime();
```

```
while(1) {  
    poll(0,0,1); /* wait for clock processing  
to occur while we're asleep*/  
    s = gethrtime();  
    while ((gethrtime() - s)/1000000 < 9);  
/* loop until almost next tick */  
}
```

==> real CPU usage: 90%

==> under Solaris9: 0% (!)

CPU/ Core/ Threads/ CMT Idea

- CPU – independent, mostly scalable
- Core – mostly / near the same
- Thread – runs on Core (!)
- T2000: 1 CPU -> 8 Cores -> 32 Threads
 - > How many **truly simultaneous** tasks may we run on it?
 - > Will it work better comparing to the classic 8CPU model?...
 - > Comparing to 16CPU model?...
 - > If Core is looping in wait -> give “hands” to another Thread!
 - > Also, think about “Threads vs Processes”

Most “popular” CMT misunderstands...

- NOTE: all items are TRUE stories!...
- T2000 is a 32CPU server
- If I disable 3 threads per Core and leave only one thread “enabled” alone - it'll run 4 times faster!...
- CPU% Busy is not proportional to my workload!...
- My batch runs faster on my laptop vs T2000!...
- Database on internal disks runs slowly!...
- etc...

Some true “real” numbers

- T2000 8cores 1200Mhz = V890 8cores 1800Mhz
 - > Oracle OLTP intensive workload
- SAP Migration (+network latency)
 - > V440 (4x900Mhz)=> 1500 SAPS
 - > T2000 (1Ghz) => 5000 CAPS
- T5120 speed-up web-oriented applications
 - > on chip chrypto
 - > 10Gbit
 - > FP is not a killer anymore :-)
- etc...

CPU / Core / Thread / ...

CPU Load Test @24CPU SPARC64-VI (48 cores, bi-thread)

Test "Data in CPU Cache"

Parallel Tasks	Time Thread-ON	Time Thread-OFF
1x	95s	95s
24x (0.5x)	102s	95s
48x (1.0x)	119s	98s
96x (2.0x)	216s	194s
192x (4.0x)	417s	382s
384x (8.0x)	819s	749s

Test "Data off Cache "

Parallel Tasks	Time Thread-ON	Time Thread-OFF
1x	39s	39s
24x (0.5x)	74s	45s
48x (1.0x)	115s	122s
96x (2.0x)	170s	226s
192x (4.0x)	327s	452s
384x (8.0x)	648s	907s

Cont. CPU / ...

CPU Load Test @20CPU SPARC64-VII (80 cores, bi-thread)

Test "Data in CPU Cache"

Parallel Tasks	Time Thread-ON	Time Thread-OFF
1x	94s	94s
20x (0.25x)	96s	98s
40x (0.5x)	96s	101s
80x (1.0x)	102s	107s
160x (2.0x)	161s	206s
320x (4.0x)	313s	396s
640x (8.0x)	611s	793s

Test "Data off Cache"

Parallel Tasks	Time Thread-ON	Time Thread-OFF
1x	38s	38s
20x (0.25x)	60s	42s
40x (0.5x)	120s	126s
80x (1.0x)	149s	154s
160x (2.0x)	180s	300s
320x (4.0x)	352s	579s
640x (8.0x)	700s	1135s

Performance: from worse to better...

- Application :-)
- Network
- I/O Subsystem & Storage
- Processes
- CPU
- ==> **Memory**



Memory Monitoring

- vmstat
 - > r w b ==> if 'w' is not null system already got RAM short
 - > sr: scan rate != 0 => problem! (since sol8)
 - > vmstat -p => page activity (MEMSTAT)
- trapstat -t
 - > TLB-miss
- MPO
 - > lgrpinfo, plgrp
- Collapsing

MPSS

- `pmap -sx PID`
- `trapstat -t 5`



Trapstat

```
# trapstat -t 5
```

cpu m	itlb-miss	%tim	itsb-miss	%tim	dtlb-miss	%tim	dtsb-miss	%tim	%tim
0 u	0	0.0	0	0.0	19	0.0	0	0.0	0.0
0 k	0	0.0	0	0.0	4	0.0	0	0.0	0.0
1 u	0	0.0	0	0.0	0	0.0	0	0.0	0.0
1 k	0	0.0	0	0.0	0	0.0	0	0.0	0.0
2 u	0	0.0	0	0.0	0	0.0	88	0.0	0.0
2 k	0	0.0	0	0.0	1373685	10.9	1	0.0	10.9
3 u	0	0.0	0	0.0	0	0.0	0	0.0	0.0
3 k	0	0.0	0	0.0	5	0.0	0	0.0	0.0

```
...
```

```
#
```


SWAP – eternal question :-)

- SWAP reservation / monitoring
 - > # swap -s
 - > # vmstat -S => swap in / swap out (!)
 - > # iostat => monitor swap device
- ISM memory
- DISM memory

Memory allocation

- System malloc()
- Multi-Threaded
 - > -l mtmalloc (Solaris)
 - > -l umem (Solaris)
 - > -l tcmalloc (Linux)
 - > -l hoard (any)
- Malloc contention
 - > # plockstat -C -p <PID>

Performance: from worse to better...

- Application :-)
- Network
- I/O Subsystem & Storage
- Processes
- CPU
- Memory
- ==> **Kernel**
- NOTE: “health altitude”...



Zones Monitoring

- vmstat per Zone
- iostat ... (hm...)
- netLOAD ... (hm...)
- prstat -Z
- psSTAT -e -o ... -M zone -T interval
- DTrace
- etc...

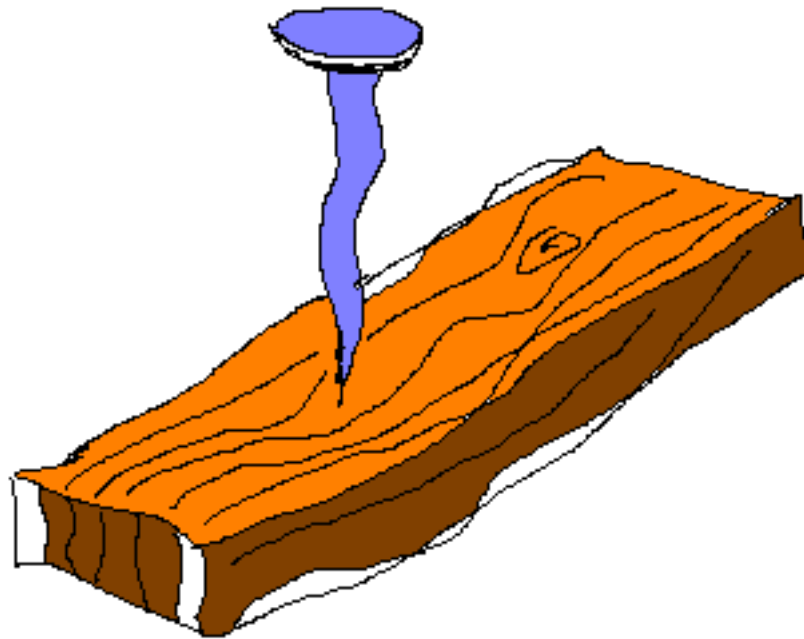
Kernel Monitoring

- lockstat
 - > lockstat -lkW sleep 10 | more
 - > lockstat -HW sleep 20 | head
- Dtrace
 - > Any danger?...
- Dtrace ToolKit
 - > hotkernel
 - > iopattern
 - > shortlived.d
 - > errinfo
 - > etc. etc. etc.

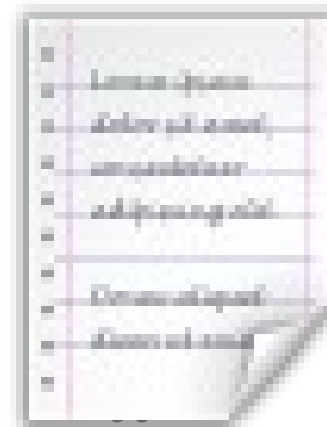
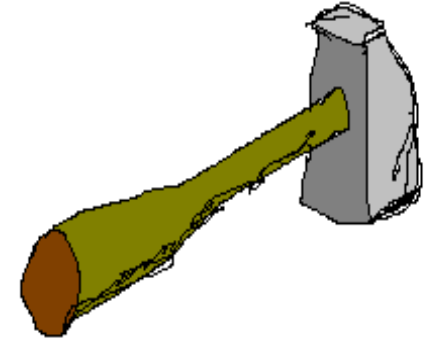
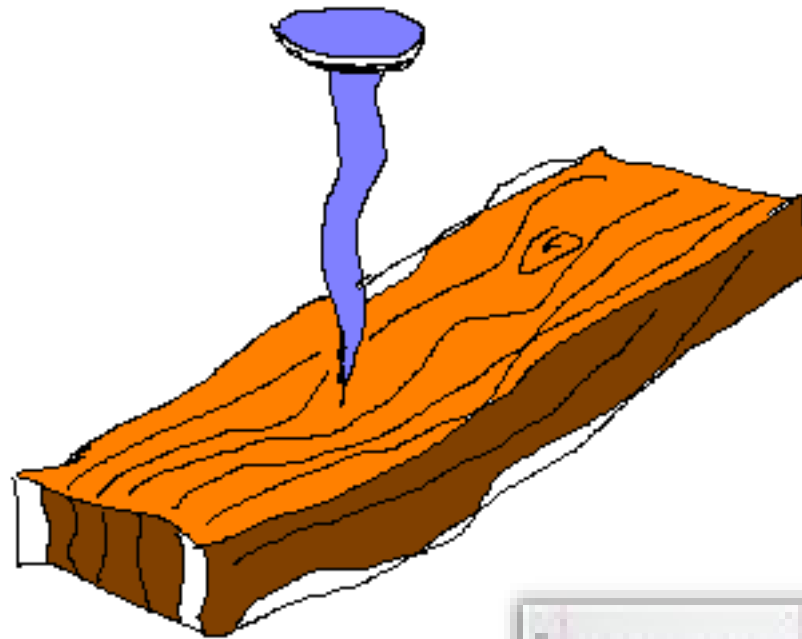
Part 2...

dim_STAT

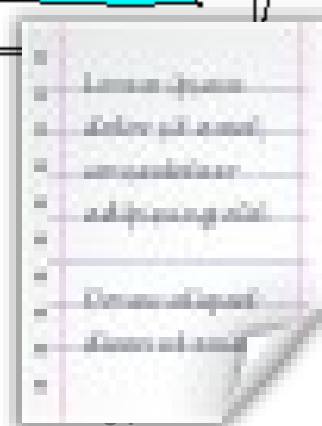
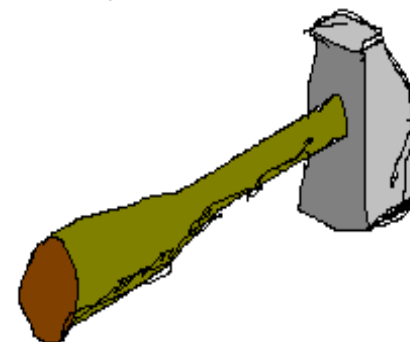
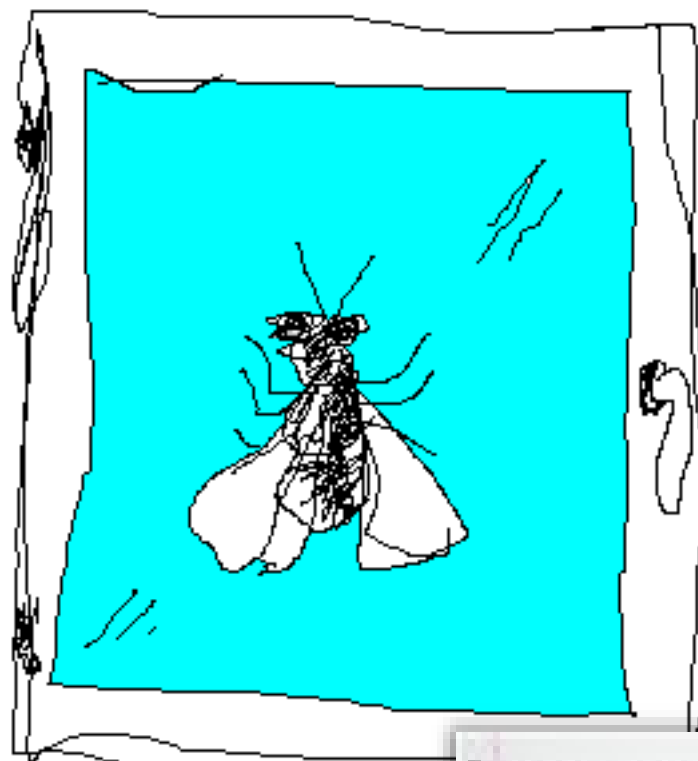
Right Tool?..



Your choice?.. :-)



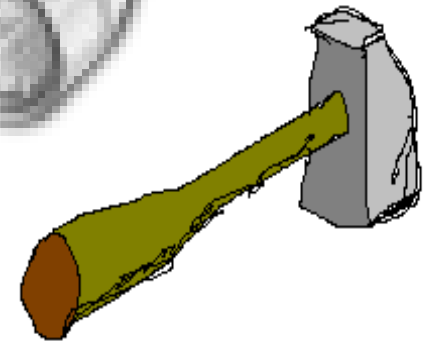
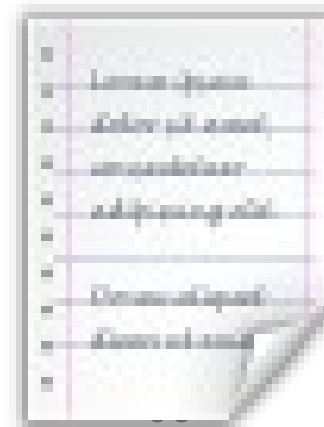
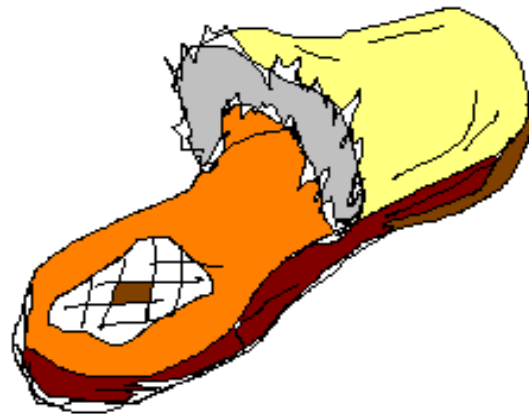
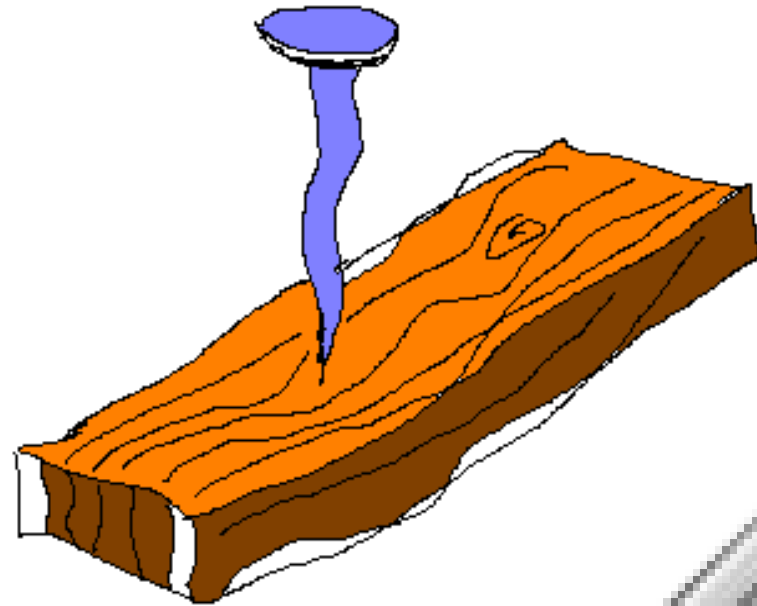
And now?.. :-)



And both?.. :-)



+



Discover Perf problem...

- Do we really have a problem?...
- Bug or feature?
- Before vs Now?
 - > sometime the only way to understand issue
 - > constant systems monitoring sees problem ahead
 - > need a small initial effort, but saves a lot
- GUDs
- EasySTAT
 - > [http:// dimitrik.free.fr/ STATsrv.pkg](http://dimitrik.free.fr/STATsrv.pkg) or [STATsrv.tar](#)
 - > (google: dim_STAT)

Part2: dim_STAT

Help yourself...

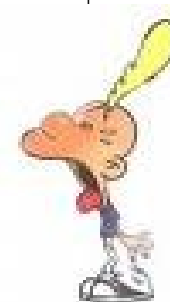
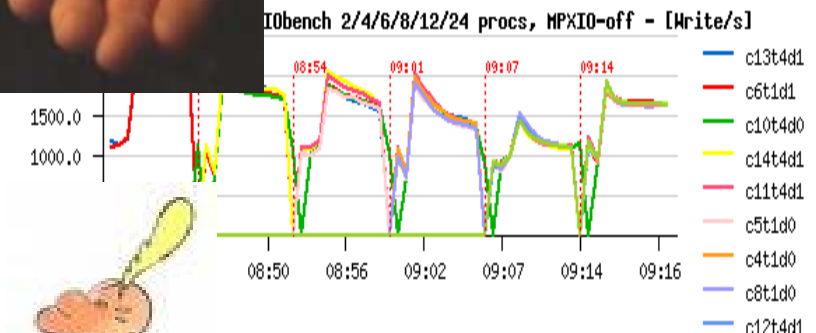
or

NO stupid work!

Why home-made tools?...

- Don't like stupid work...
- Tool adapted to Humans, and not Humans to the tool
- Best implementations when User = Developer
- Mind Pleasure
- Laziness = Power of Progress!
- Machines should help Humans!



A close-up photograph of two human hands, palms up, holding small, oval-shaped pills. The left hand holds a red pill, and the right hand holds a blue pill. The background is dark, and the lighting highlights the texture of the skin and the smooth surface of the pills.

Performance Monitoring...

- System and application stats are our friends
- Only raw data – hard! (ex: days -> weeks load view)
- Only graph data – hard! (ex: need exact numbers)
- Command line interface (CLI) is most common
- Data presentation is most painful...
- Time interval is very important for accurate measure
- Too much data = out of control...
- Too few (avg) data = out of detail (problem)...
 - > Ex: Avg(temperature) in hospital is OK



Real life example...

- Customer claims:
 - > my CPU is used only 20%
 - > my Run Queue is always 10 (!)
 - > so: Solaris has a problem!...
- How it's possible?
- Bug?
- Well, how did you see it?...

Real life example (cont.)

- \$ vmstat 5

```
> r w b ..... us sy id
10 0 0 ..... 20 0 80
10 0 0 ..... 20 0 80
10 0 0 ..... 20 0 80
10 0 0 ..... 20 0 80
10 0 0 ..... 20 0 80
....
```

- Hmm...

Real life example (cont. 2)

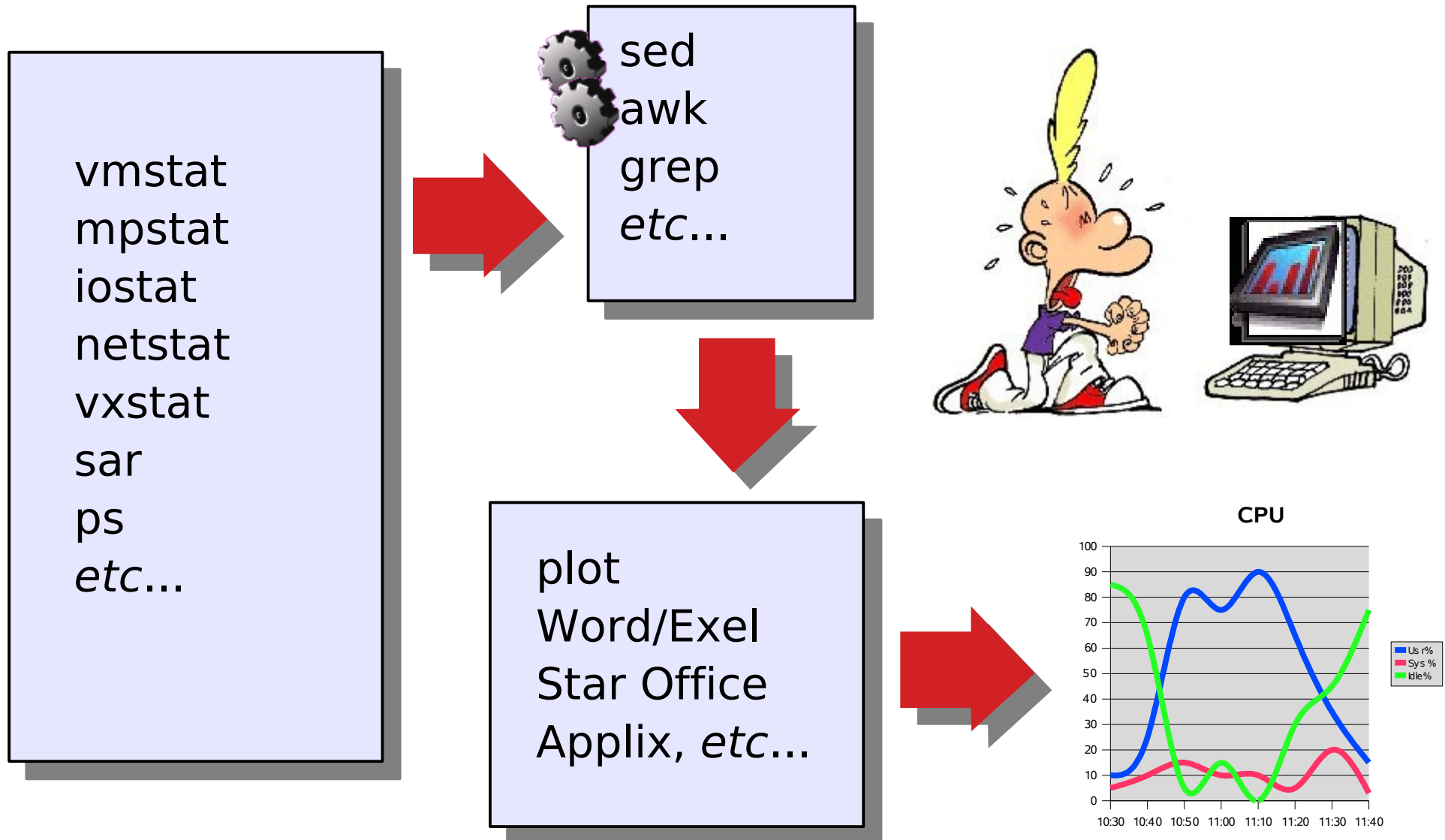
- But what if?...

- \$ vmstat 1

```
> r w b ..... us sy id
50 0 0 ..... 100 0 0  <===
0 0 0 ..... 0 0 100
0 0 0 ..... 0 0 100
0 0 0 ..... 0 0 100
0 0 0 ..... 0 0 100
50 0 0 ..... 100 0 0  <===
0 0 0 ..... 0 0 100
0 0 0 ..... 0 0 100
0 0 0 ..... 0 0 100
0 0 0 ..... 0 0 100
```

- Got it? - AVG or not AVG :-)

Benchmark Center in 1997...

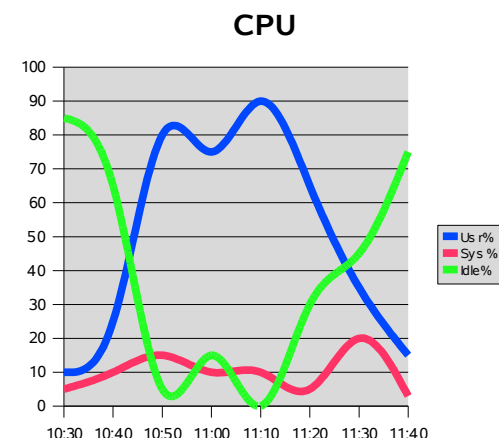


Main Idea...

vmstat
mpstat
iostat
netstat
vxstat
sar
har
psSTAT
etc...

**Real Time
Timestamped
Correlated**

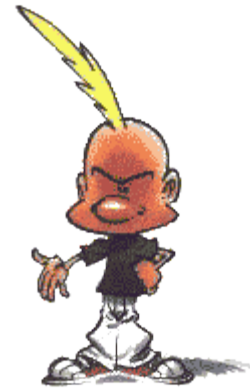
**SQL
Database**



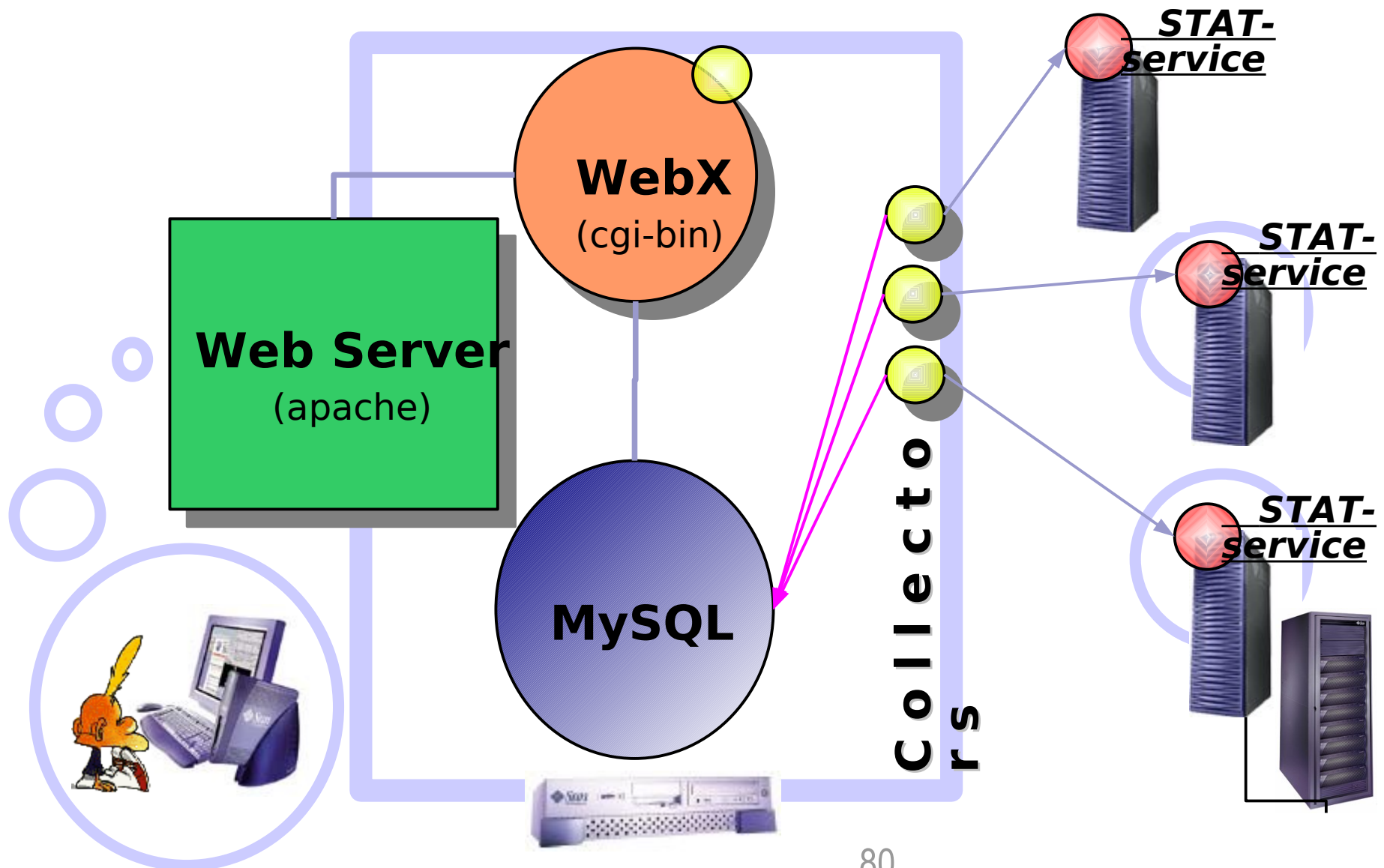
JDBC, ODBC, Native SQL, Pro*C, etc...
PHP, Perl, Tcl, Python, WebX, etc...
Word/Excel, StarOffice, Applix, etc...
etc...

Wish list for “STAT” tool

- Run on Solaris! :-)
- Really small overhead!
- Easy to install
- Easy to use & administrate
- Easy to extend
- Easy to access internal data
- Unlimited history
- Application feed-back
- Without external cost - Why should we pay to analyze our own machines?...



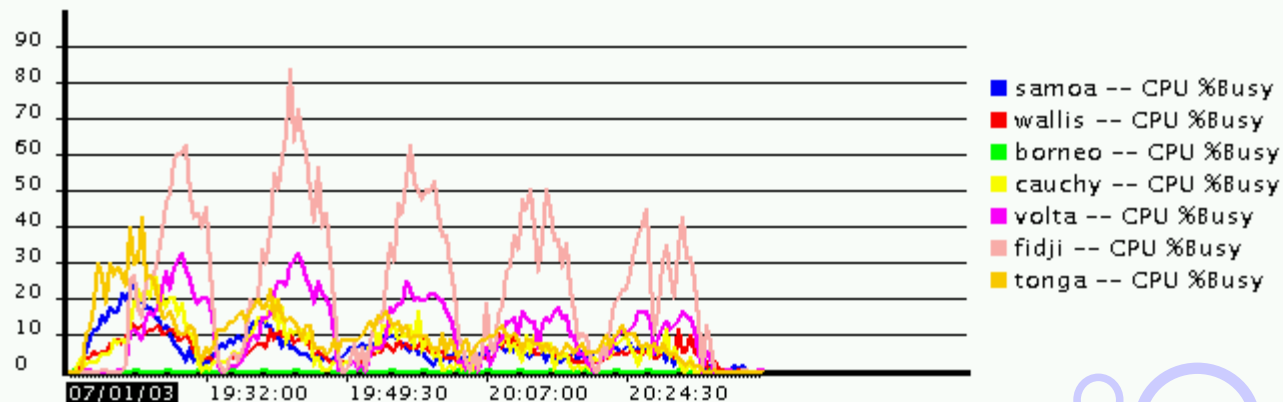
dim_STAT Architecture Overview



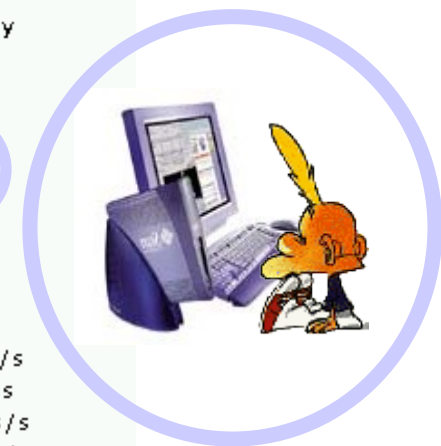
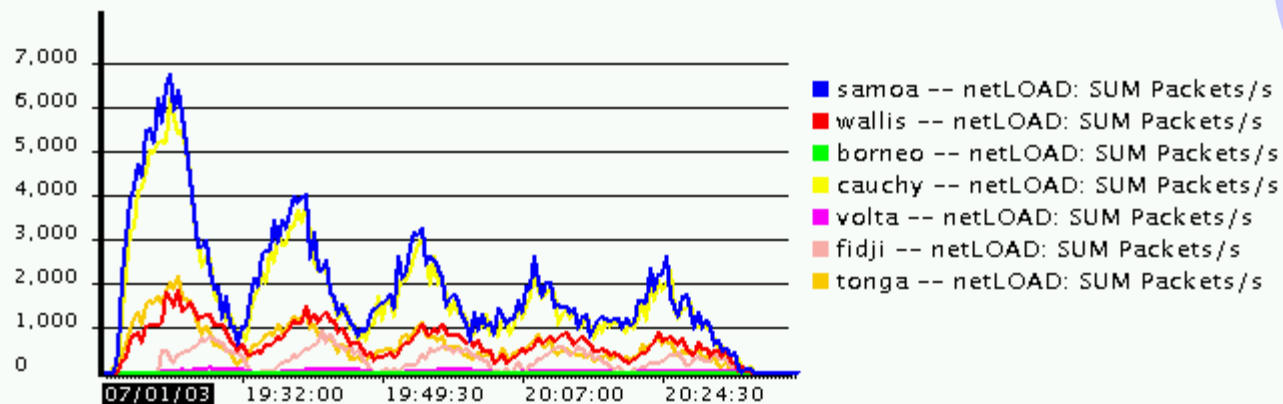
Example Multi-host Analyzing...

dim_STAT Multi-Host Analyzer

CPU %Busy



netLOAD: SUM Packets/s



STAT-service details

- Controlled & logged access
- On-Demand Start/Stop service
- Listening on TCP/IP port and publishing available stats
- Includes command line kit (**EasySTAT**) to collect data locally and load later (paranoid or very protected sites)
- Default:
 - > *vmstat, mpstat, iostat, netLOAD, ProcLOAD, UserLOAD, vxstat, **ZoneLOAD**, PoolLOAD, ProjLOAD*
- Extended:
 - > *jvmSTAT (**JVM** mem.usage, GC activity)*
 - > **Oracle:** *oraEXEC, oraIO, oraENQ, **MySQL**, PostgreSQL*
- and any New you want to add! (Note: GPL since v.8.1)



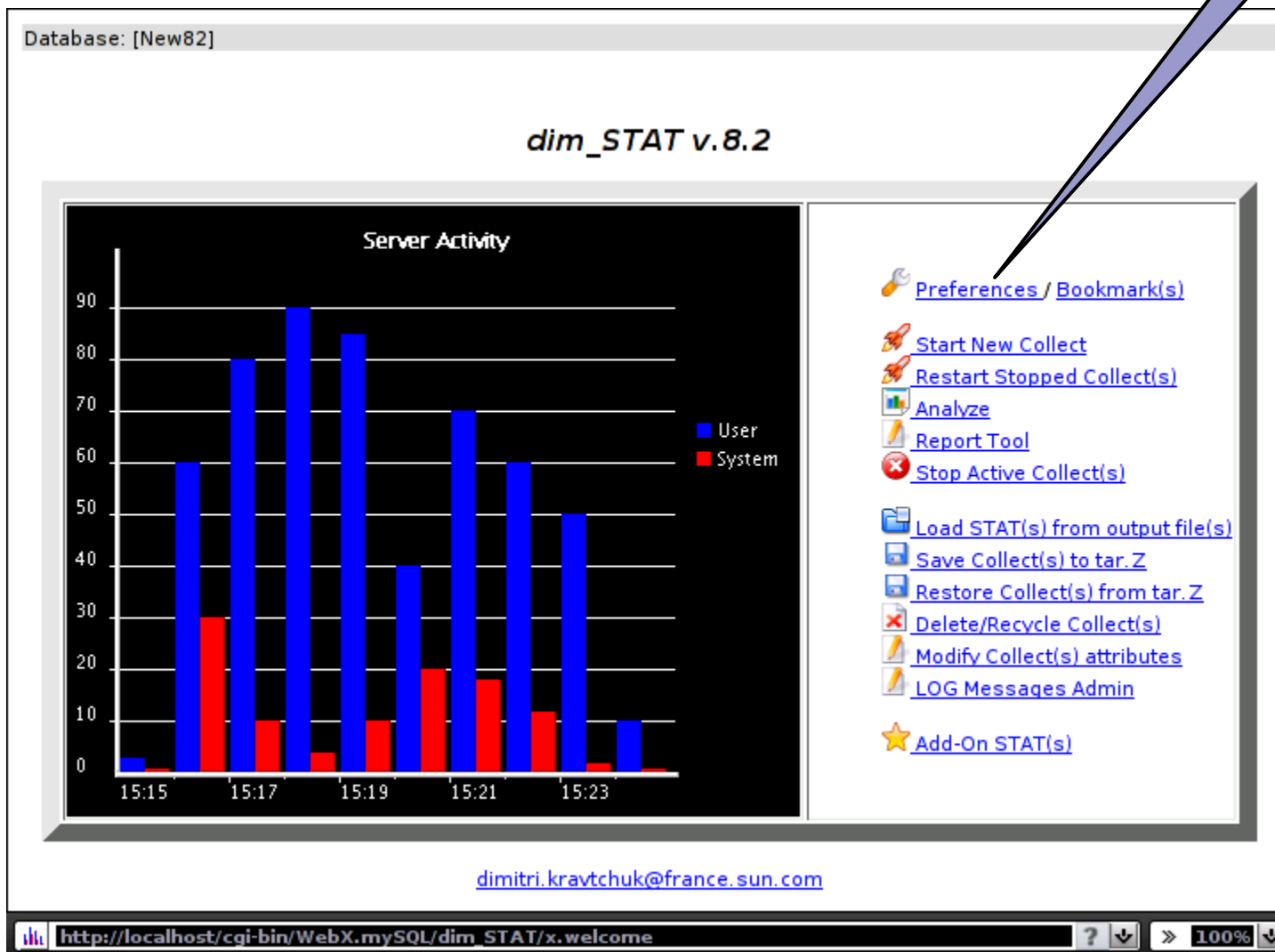
STAT-service Access Config

- # pkgadd -d STATsrv.pkg
- # /etc/STATsrv/STAT-service start
- /etc/STATsrv/log/access.log <== service logging
- /etc/STATsrv/access <== Main config file

```
# .....  
# // All following commands should work out of the box...  //  
# .....  
  
command  vmstat          /usr/bin/vmstat  
command  mpstat          /usr/bin/mpstat  
command  netstat         /usr/bin/netstat  
command  ForkExec        /etc/STATsrv/bin/ForkExec.sh  
command  MEMSTAT         /etc/STATsrv/bin/MemStat.sh  
command  tailX           /etc/STATsrv/bin/tailX  
command  ioSTAT.sh       /etc/STATsrv/bin/ioSTAT.sh  
...  
```

Main Page

Go!



Preferences setting...

Database: [New82] [\[Home \]](#) [\[Preferences \]](#) [\[LOG Admin \]](#)

dim_STAT Preferences

Database

Database

New Database: **Create!**

Use Database: **New82**

Current Database: **New82**

Used Space: 3 MB

Biggest File: 1 MB

Free Space: 467 MB

Host names

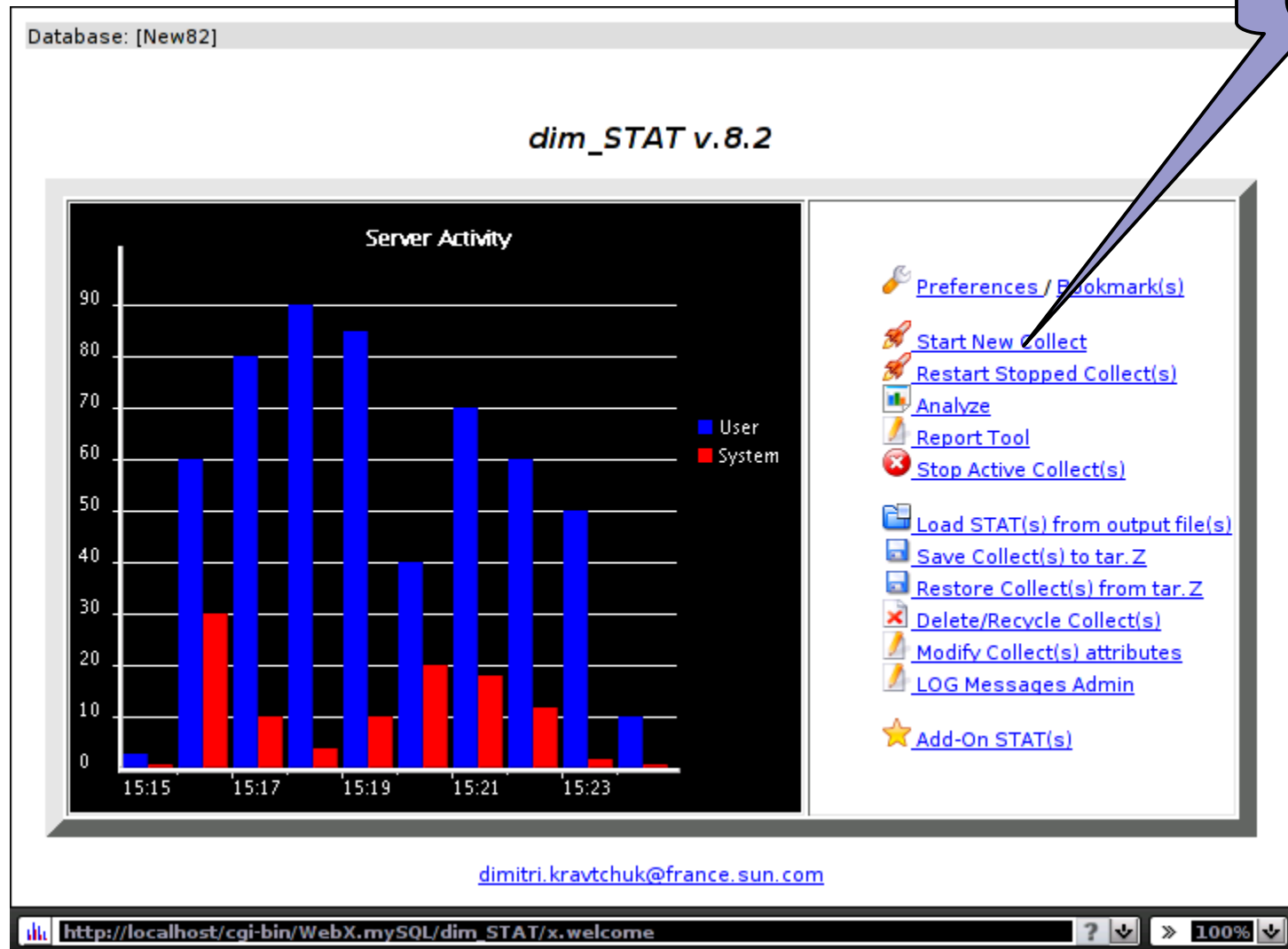
Host Name List

```
dimitri:5000
localhost
v890/LLG-free
v490/laplace-llg
uk-server/10.132.0.243
```

http://dimitri/cgi-bin/WebX.mysql/dim_STAT/x.preferences ? 100%



Start New Collect



Go!








Start new STAT collect...

Database: [New82]
[\[Home \]](#)
[\[Preferences \]](#)
[\[LOG Admin \]](#)

Select hosts




Host(s)


☐  dimitri:5000
☐  localhost
☐  v890/LLG-free
☐  v490/laplace-llg
☐  uk-server/10.132.0.243
☐ [ALL]

New:

☐ Preselect all STATs matching:

LED Description:

-  - ok: STAT-service is running on the host, compatible version
-  - degraded: STAT-service is running on the host, but old version, please, upgrade!
-  - bad: no STAT-service running or host is not accessible...
- NOTE: your default STAT-service port is 5000


http://dimitri/cgi-bin/WebX.mysql/dim_STAT/x.client.pll
?
100%



Start new STAT collect (cont.)

Select stats




Database: [New82] [\[Home \]](#) [\[Preferences \]](#) [\[LOG Admin \]](#)

dim_STAT Start New Collect(s)

Collect BaseName:

Stat Title:

Time Interval: sec.

Host	Stat ID	STAT(s)	Client Log Filename
 dimitri	<input type="text" value="6"/>	<input type="checkbox"/> [*] <input type="checkbox"/> IObench <input type="checkbox"/> LcpuSTAT <input type="checkbox"/> LioSTAT <input type="checkbox"/> LnetLOAD <input type="checkbox"/> LPrclOAD <input type="checkbox"/> LpsSTAT <input type="checkbox"/> LUsrLOAD <input type="checkbox"/> Lvmstat	<input type="text"/>
 v890	<input type="text" value="7"/>	<input type="checkbox"/> [*] <input type="checkbox"/> VMSTAT <input type="checkbox"/> MPSTAT <input type="checkbox"/> IOSTAT <input type="button" value="Active-ONLY"/> <input type="checkbox"/> NETSTAT <input type="checkbox"/> psSTAT <input type="button" value="Active-ONLY"/> <input type="checkbox"/> ForkExec <input type="checkbox"/> MEMSTAT <input type="checkbox"/> netLOAD <input type="checkbox"/> ProcLOAD <input type="checkbox"/> UserLOAD	<input type="text"/>
 uk-server	<input type="text" value="8"/>		

☐ Show Debug Output

WARNING: ... STAT ...

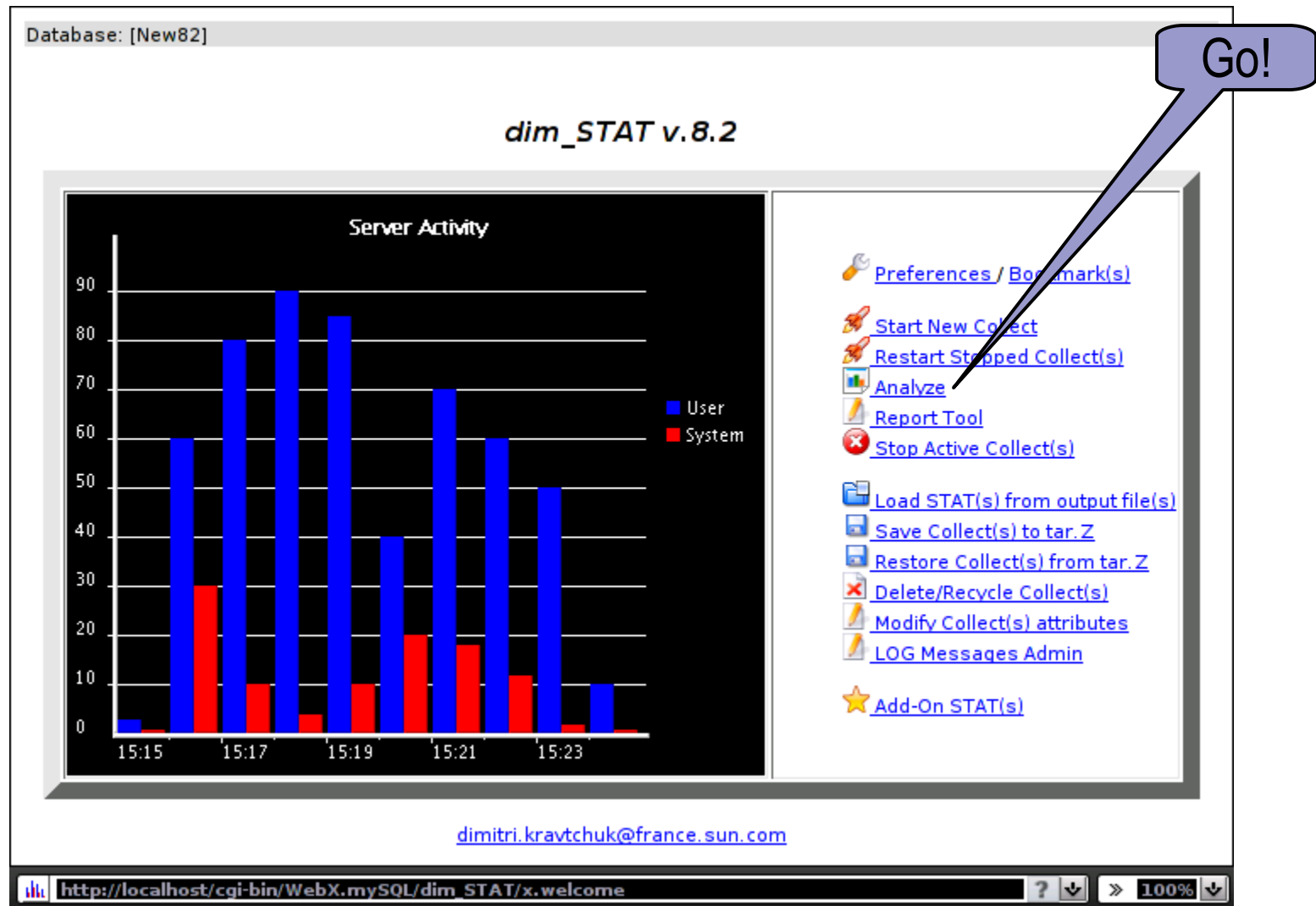
http://dimitri/cgi-bin/WebX.mySQL



BatchLOAD, EasySTAT

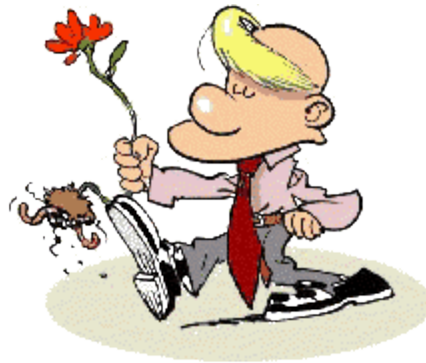
- BatchLOAD
 - > universal stats loader from flat files (guds, other)
- EasySTAT
 - > integrated into STAT-service
 - > # nohup /etc/STATsrv/bin/EasySTAT.sh /var/stats 30 24 &
 - > # gtar czf stats.tgz /var/stats; mailto stats.tgz
 - > \$ gtar xzvf stats.tgz
 - > cd /var/stats; vi LoadDATA.sh <== adapt params
 - > sh LoadDATA.sh <== load
 - > Analyze! :-)

Analyze



Analyzer interface

- Single or Multi-Host view
 - > Single: more detailed, in depth analyze
 - > Multi: global view, relative to hosts activity
- Choose Host(s)
- Choose time period
- Choose STATs and presentation mode (text, graph)
- Go!



Example: Multi-Host Analyze

Database: [New82]
[\[Home \]](#)
[\[Preferences \]](#)
[\[LOG Admin \]](#)

dim_STAT Analyze

Multi host

☐ Single-Host Analyze
☒ Multi-Host Analyze

☐ Active ONLY
☐ Preselect Multi-Hosts
☐ Show STAT(s) Status
Titles Matching: *
Hostname Matching: top*
LOG Messages Matching: *

hosts, title pattern

Analyze Reset

http://dimitri/cgi-bin/WebX.mysql/dim_STAT/x.welcome_analyze
100%



Example: Multi-Host Analyze (cont.)

Database: [New82] [\[Home \]](#) [\[Preferences \]](#) [\[LOG Admin \]](#)

dim_STAT Multi-Host Analyzer

	ID	Host	Title	Started	Interval	State
<input type="checkbox"/>	3	top-A	Probe new Multi-host (@10.128.4.70)	2007-07-06 21:59:19	15	Finished
<input type="checkbox"/>	4	top-C	Probe new Multi-host (@10.128.4.74)	2007-07-06 21:59:22	15	Finished

☐ Check/Uncheck All

☒ All data
☐ Last minutes (current time)
☐ Between and DateTimes YYYY-MM-DD HH:MI
☐ After LOG Message minutes
☐ Between LOG Messages

Interval

```
[22:16 2007.07.06] top-A: STOP!
[22:16 2007.07.06] top-C: STOP!
[22:16 2007.07.06] top-A: STOP!
[22:16 2007.07.06] top-C: STOP!
[22:16 2007.07.06] top-A: STOP!
[22:16 2007.07.06] top-C: STOP!
[22:16 2007.07.06] top-A: STOP!
[22:16 2007.07.06] top-C: STOP!
```

☐ AVG value(s) by every measurement(s)

http://dimitri/cgi-bin/WebX.mysql ? 100%

Time period



Example: Multi-Host Analyze (cont. 2)

Select stats

Values

[22:16 2007.07.06] top-A: STOP!

[22:16 2007.07.06] top-C: STOP!

☐ AVG value(s) by every measurement(s)

☒ Per host
 ☐ Grouped AVG by first/last letters in host name
 ☐ Grouped SUM by first/last letters in host name
 ☐ Grouped MAX by first/last letters in host name
 ☐ Grouped MIN by first/last letters in host name

Note: if value already contains SUM/AVG/etc. grouping, host's aggregate will be ignored

Graphics

Mode: Style:

Background: Size: x

☐ Force Graphs alignment
 ☐ Force Data Gap Completion

Title:

Values:

☐ CPU: Ustr%
 ☐ CPU: Sys%
 ☒ CPU: Busy%
 ☐ RAM: Free List (KB)
 ☐ Run queue
 ☐ Blocked processes
 ☐ I/O: Read KB/s

☐ netLOAD: SUM O_errors/s
 ☐ oraENQ: SUM ReqERR
 ☐ oraENQ: SUM ReqOK
 ☐ oraENQ: SUM ReqTot
 ☐ oraENQ: SUM Waits
 ☐ oraENQ: SUM WaitTM
 ☐ oraEXEC: Commit/sec
 ☒ oraEXEC: Exec/sec
 ☐ oraEXEC: Sessions
 ☐ ProcLOAD: SUM N_activ
 ☐ ProcLOAD: SUM N_lwp
 ☐ ProcLOAD: SUM N_total

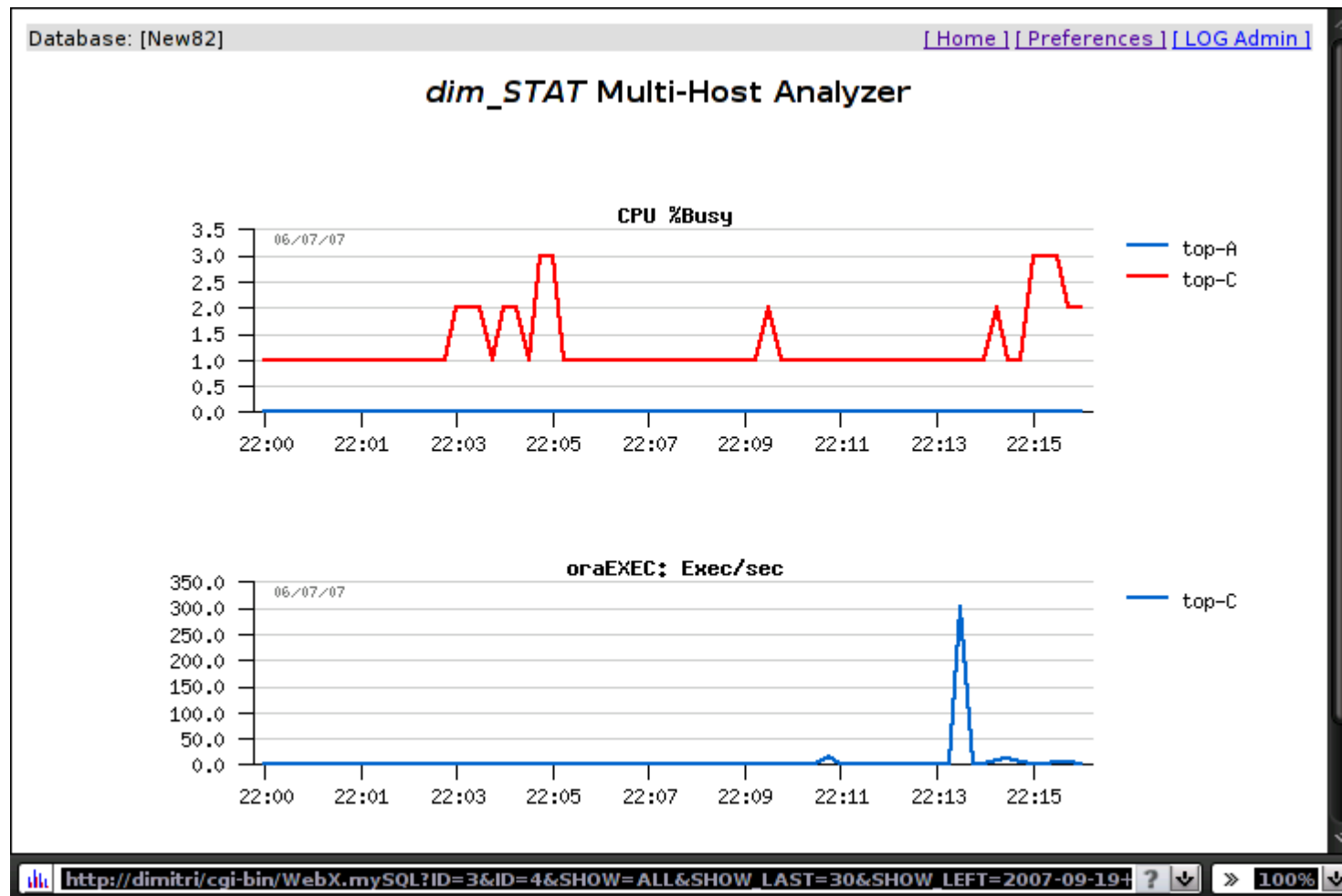
☐ Check/Uncheck All

☐ Show Host's LOG Messages Matching:
☐ Show Task(s) from Host(s)

☐ Refresh every sec.

http://dimitri/cgi-bin/WebX.mysql

Example: Multi-Host Analyze (cont.3)



Example: Single-Host Analyze

Database: [New82] [\[Home \]](#) [\[Preferences \]](#) [\[LOG Admin \]](#)

dim_STAT Analyzer

	ID	Host	Title	Started	Interval	State
<input type="radio"/>	3	top-A	Probe new Multi-host (@10.128.4.70)	2007-07-06 21:59:19	15	Finished
<input checked="" type="radio"/>	4	top-C	Probe new Multi-host (@10.128.4.74)	2007-07-06 21:59:22	15	Finished

Use LogFile Messages from: **the same host** Matching: *

http://dimitri/cgi-bin/WebX.mySQL

Select host,
click on STAT



Example: Single-Host Analyze

(cont.)

Setup criteria

Database: [New82] [\[Home \]](#) [\[Preferences \]](#) [\[LOG Admin \]](#)

dim_STAT Analyzer

Solaris Disk I/O Activity Statistics (iostat) Started: 2007-07-06 21:59:22 Timeout: 15 sec.

Disk(s)

<input type="checkbox"/> c0t10d0	<input type="checkbox"/> c14t60x6xB9C1d0
<input type="checkbox"/> c0t11d0	<input type="checkbox"/> c14t60x7xC190d0
<input type="checkbox"/> c0t8d0	<input type="checkbox"/> c14t60x8xB75Fd0
<input type="checkbox"/> c0t9d0	<input type="checkbox"/> c14t60x9xC137d0
<input type="checkbox"/> c14t60x10xBA0Cd0	<input type="checkbox"/> nfs1
<input type="checkbox"/> c14t60x11xB3FFd0	<input type="checkbox"/> nfs2
<input type="checkbox"/> c14t60x1xB6D4d0	
<input type="checkbox"/> c14t60x2xB43Ed0	
<input type="checkbox"/> c14t60x3xB7C4d0	
<input type="checkbox"/> c14t60x5xB685d0	

☐ Inversed Selection

☒ Select TOP **DISK(s)** with **MAX-Total** **Read/s**

☐ Accept ONLY data with **Read/s** **=**

☐ Use Select Pattern * for **DISK**

☐ All data

☒ Last measurements

☐ Top value(s)

<input type="checkbox"/> Read/s
<input type="checkbox"/> Read KB/s
<input type="checkbox"/> Write/s
<input type="checkbox"/> Write KB/s
<input type="checkbox"/> Wait

☐ Table of results

<input type="checkbox"/> ID
<input type="checkbox"/> Sno#
<input type="checkbox"/> DISK
<input type="checkbox"/> Read/s
<input type="checkbox"/> Read KB/s

☒ Graphics

Mode: **PNG Image** Style: **Bold ContGraph**

Background: **White** Size: x

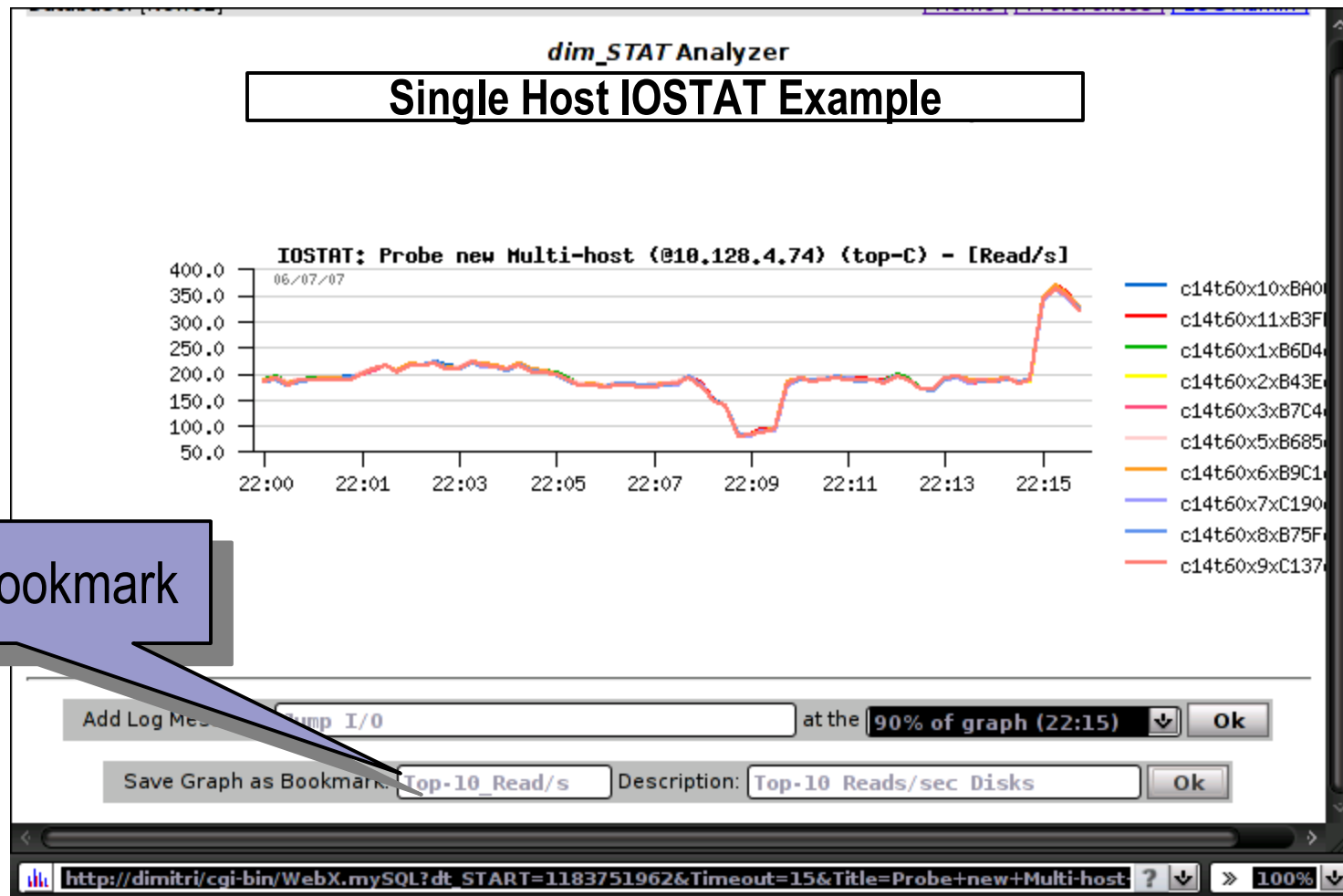
☐ Force Graphs alignment

☐ Force Data Gap Completion

http://dimitri/cgi-bin/WebX.mysql



Example: Single-Host Analyze (cont2.)

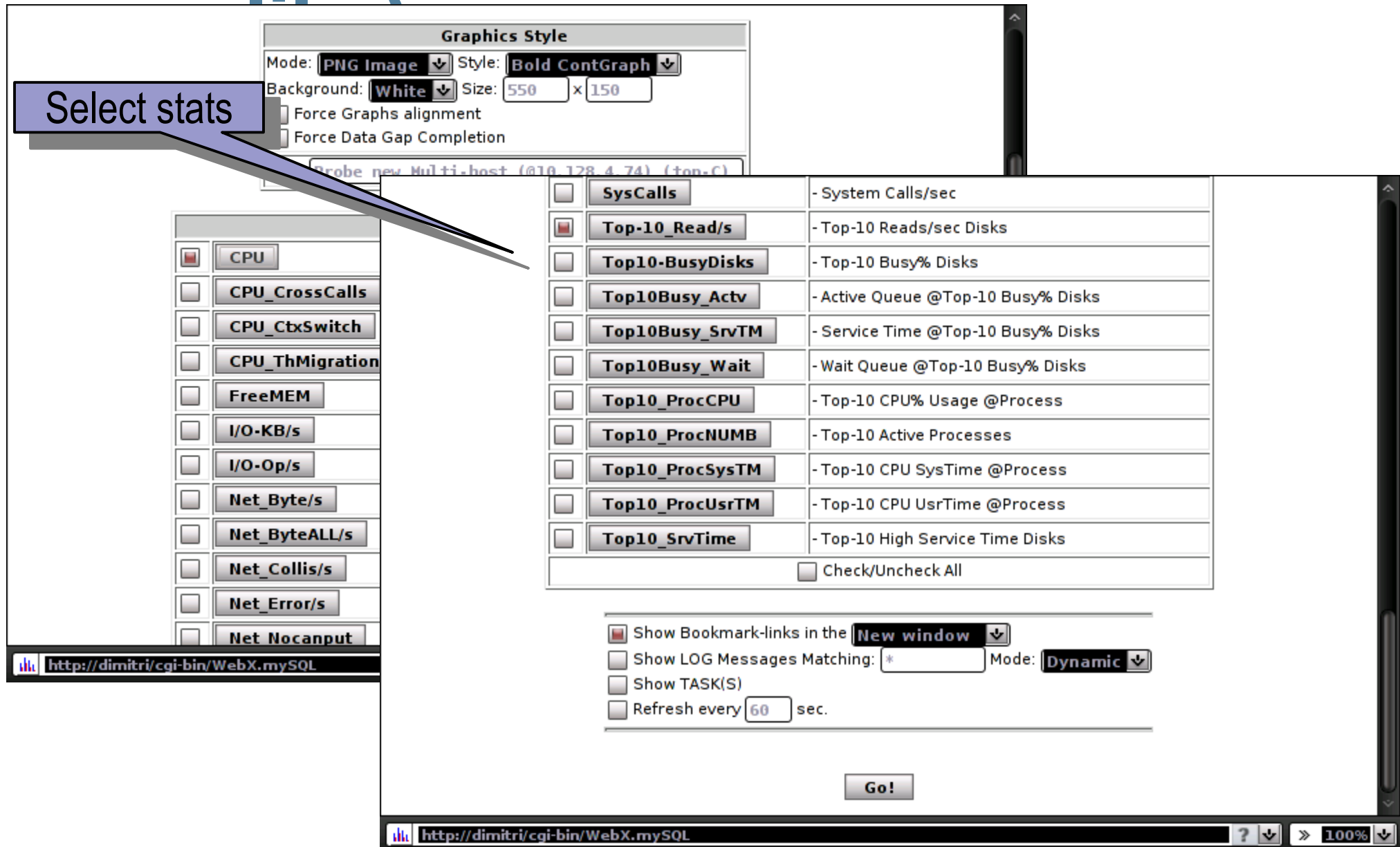


Save as Bookmark



Bookmarks (or rename-it-as-

Select stats



Graphics Style

Mode: **PNG Image** Style: **Bold ContGraph**

Background: **White** Size: **550** x **150**

☐ Force Graphs alignment

☐ Force Data Gap Completion

Probe new Multi-host (@10.128.4.74) (top-C)

<input type="checkbox"/>	SysCalls	- System Calls/sec
<input checked="" type="checkbox"/>	Top-10_Read/s	- Top-10 Reads/sec Disks
<input type="checkbox"/>	Top10-BusyDisks	- Top-10 Busy% Disks
<input type="checkbox"/>	Top10Busy_Actv	- Active Queue @Top-10 Busy% Disks
<input type="checkbox"/>	Top10Busy_SrvTM	- Service Time @Top-10 Busy% Disks
<input type="checkbox"/>	Top10Busy_Wait	- Wait Queue @Top-10 Busy% Disks
<input type="checkbox"/>	Top10_ProcCPU	- Top-10 CPU% Usage @Process
<input type="checkbox"/>	Top10_ProcNUMB	- Top-10 Active Processes
<input type="checkbox"/>	Top10_ProcSysTM	- Top-10 CPU SysTime @Process
<input type="checkbox"/>	Top10_ProcUsrTM	- Top-10 CPU UsrTime @Process
<input type="checkbox"/>	Top10_SrvTime	- Top-10 High Service Time Disks
		<input type="checkbox"/> Check/Uncheck All

☒ Show Bookmark-links in the **New window**

☐ Show LOG Messages Matching: * Mode: **Dynamic**

☐ Show TASK(S)

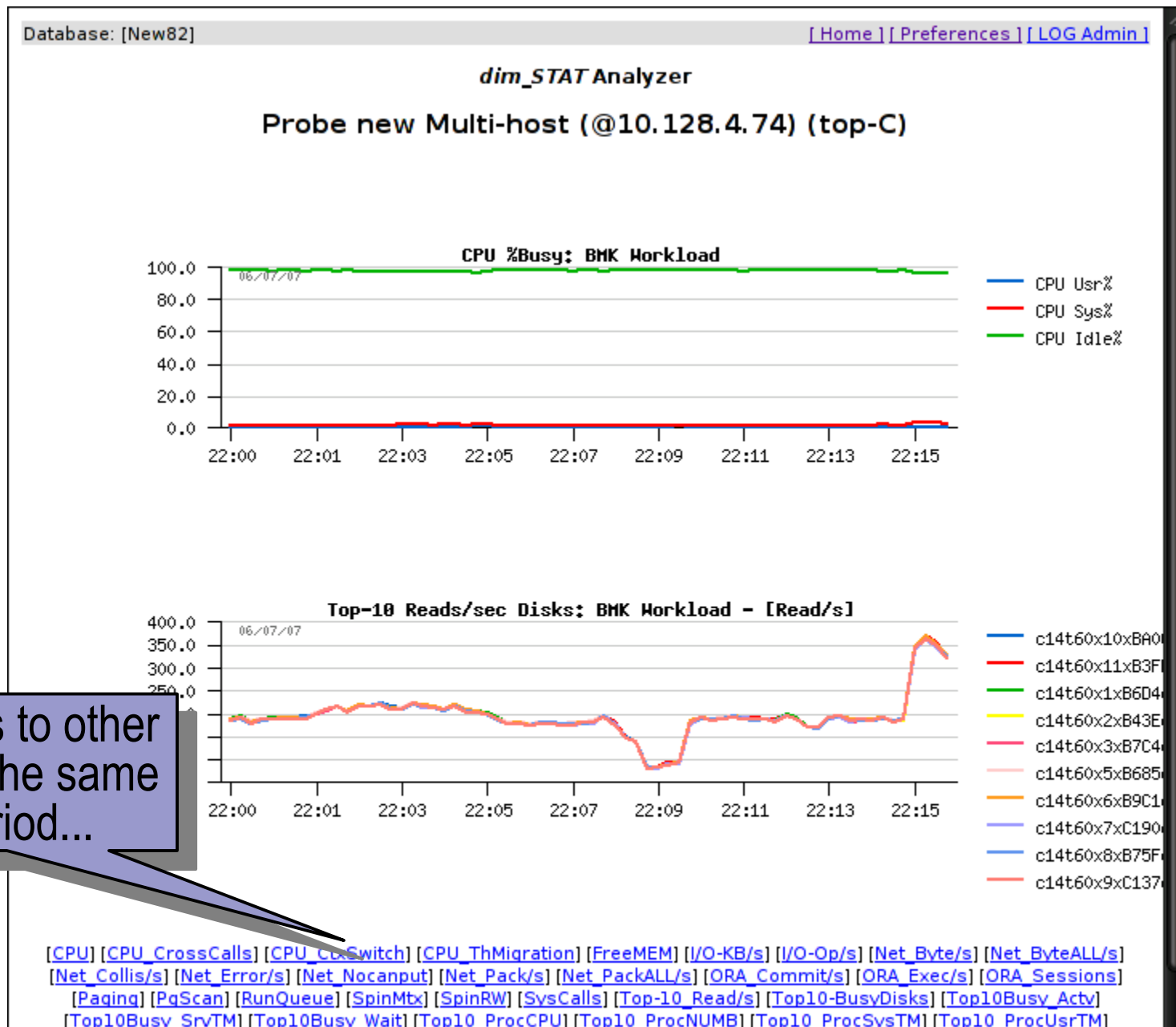
☐ Refresh every **60** sec.

Go!

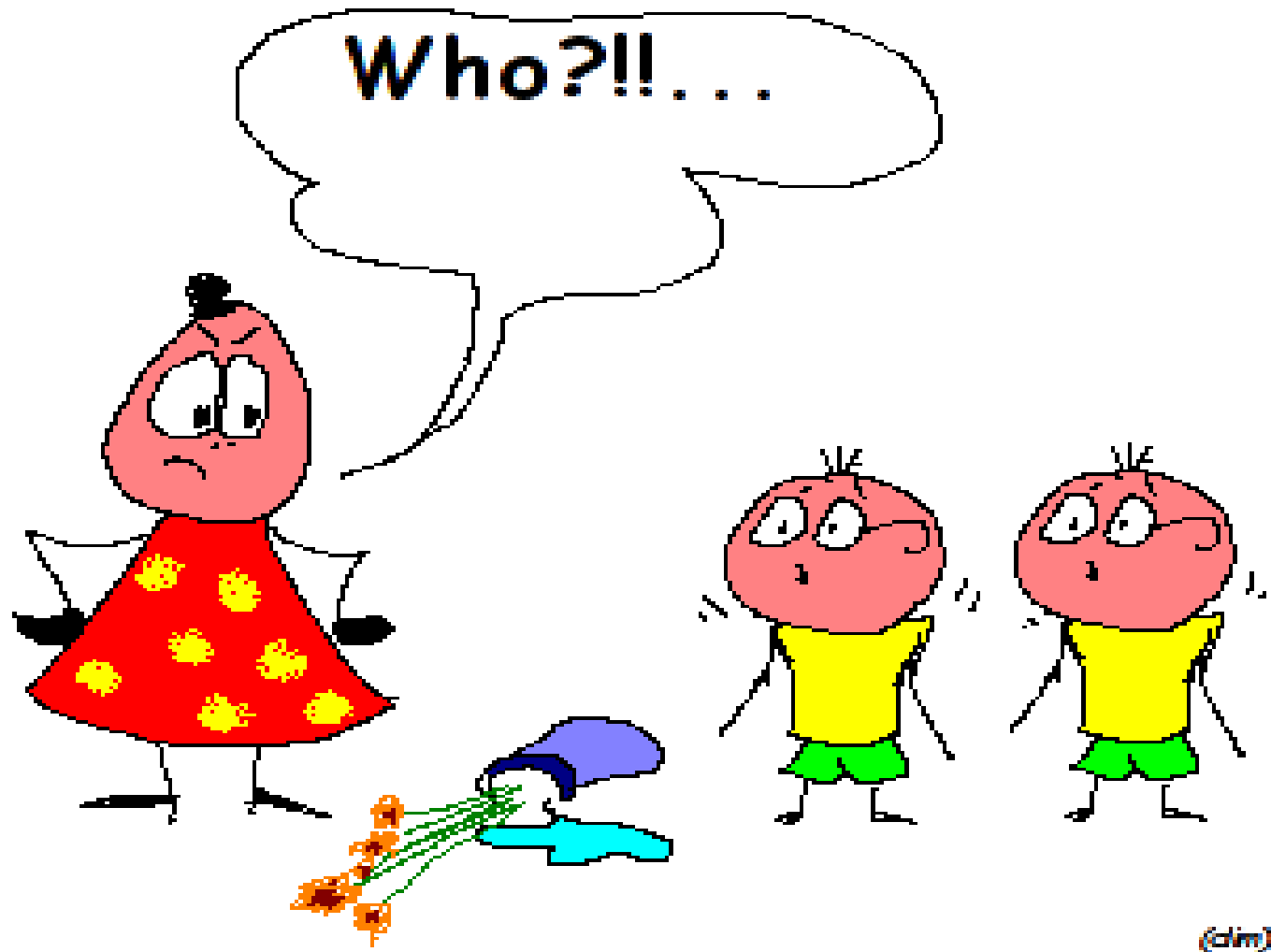
http://dimitri/cgi-bin/WebX.mysql

http://dimitri/cgi-bin/WebX.mysql

Bookmarks (cont.)

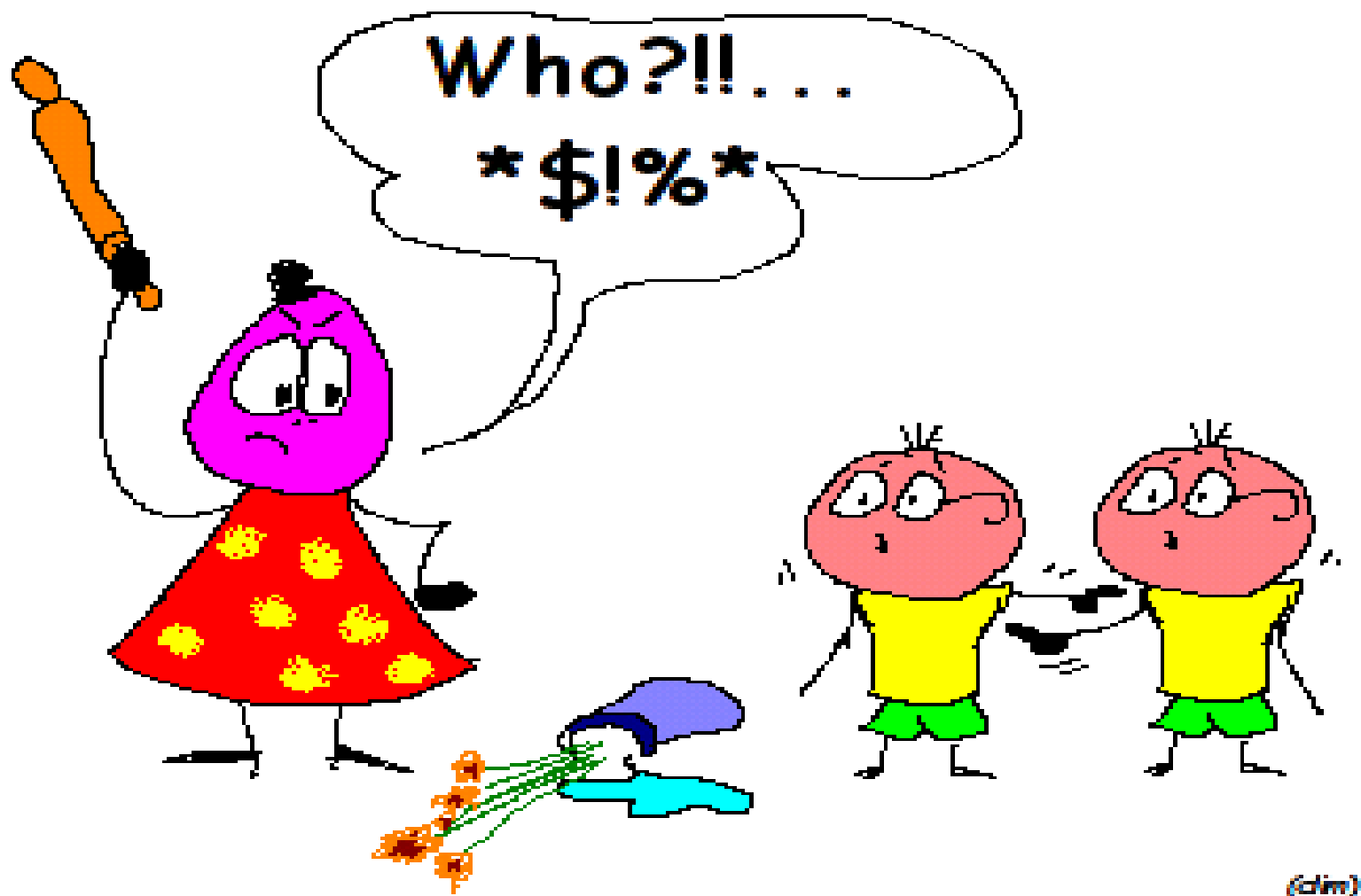


Classic case: On customer's site



(dim)

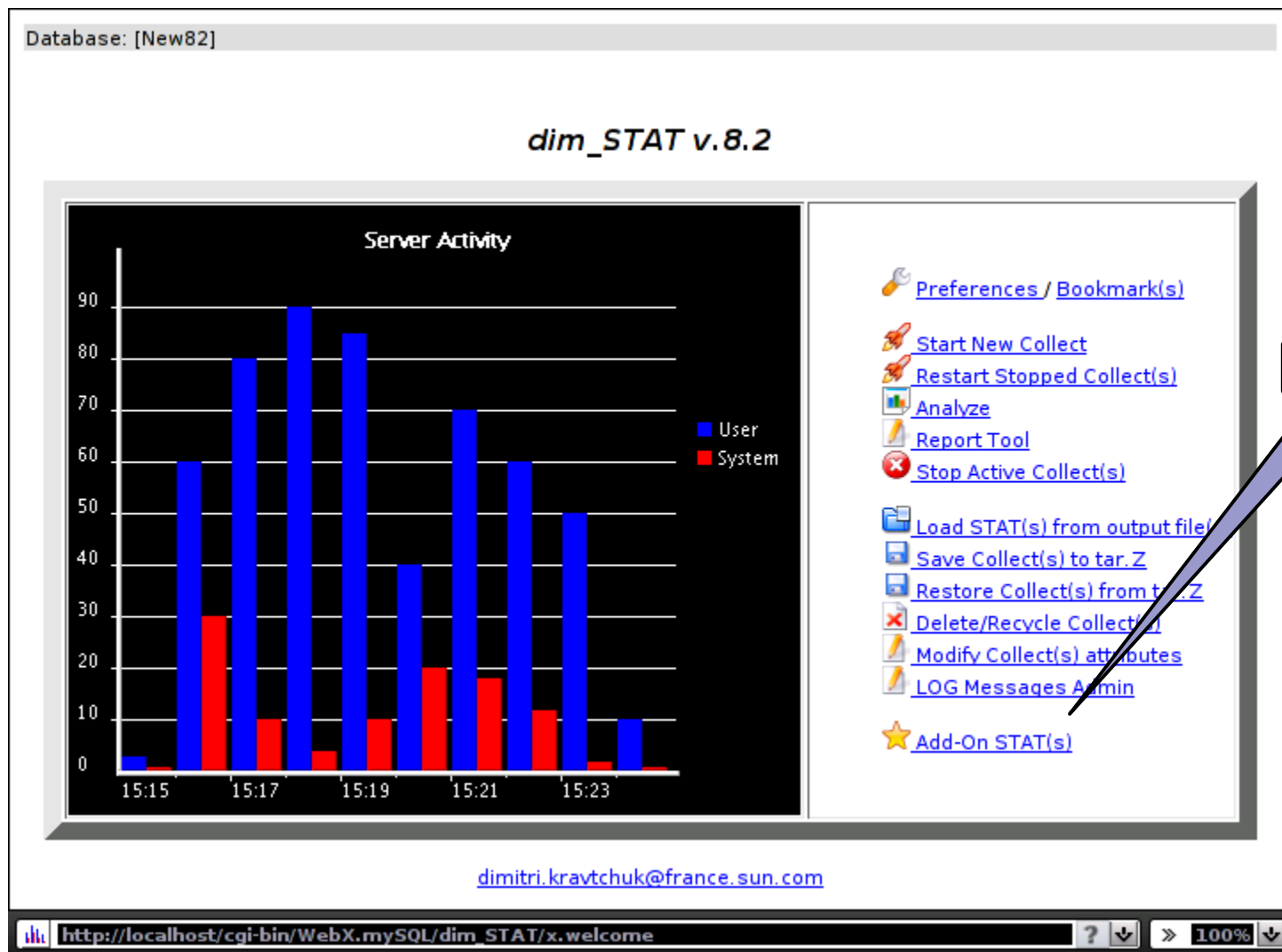
Classic case: On customer's site



EasySTAT kit

- Part of STAT-service
- Simple to automate!
 - > # /etc/STATsrv/bin/EasySTAT.sh /var/tmp/stats 15 24
 - > # cd /var/tmp; tar czf /tmp/stats-`hostname`.tgz stats
 - > # send ... /tmp/stats.tgz; rm /tmp/stats.tgz
 - > # get stats.tgz; tar xzf stats.tgz; cd stats
 - > # vi LoadDATA.sh (if needed)
 - > # sh LoadDATA.sh
- Analyze! :-)

Add-Ons



Go!



Add-Ons: new stats integration

Database: [New82]
[\[Home \]](#)
[\[Preferences \]](#)
[\[LOG Admin \]](#)

dim_STAT - Add-On STAT(s)

Current database: [New82]

[★ Integrate New Add-On-STAT](#)
[Save Add-On STAT\(s\) Description](#)
[Restore Add-On STAT\(s\) Description](#)
[Delete Add-On STAT\(s\)](#)

Add-On STAT(s) -- Standard Kit

ForkExec	-- ForkExec Statistic(s)	-[Ok]-
HPiostat	-- HP/UX iostat Statistic(s)	-[Ok]-
HPvmstat	-- HP/UX vmstat Statistic(s)	-[Ok]-
IObench	-- IObench Statistic(s)	-[Ok]-
LPrcLOAD	-- Linux ProcLOAD	-[Ok]-
LUsrLOAD	-- Linux UserLOAD	-[Ok]-
LcpuSTAT	-- Linux MPSTAT (per CPU) Statistic(s)	-[Ok]-
LioSTAT	-- Linux I/O Statistic(s)	-[Ok]-
LnetLOAD	-- Linux netLOAD	-[Ok]-
LpsSTAT	-- Linux psSTAT	-[Ok]-
Lvmstat	-- Linux VMSTAT Statistic(s)	-[Ok]-
MEMSTAT	-- Solaris Memory Related Statistics (vmstat -p)	-[Ok]-
PoolLOAD	-- PoolLOAD - Solaris 10 per Pool Load Statistic(s)	-[Ok]-
ProcLOAD	-- ProcLOAD - per Process Name Load Statistic(s)	-[Ok]-
ProjLOAD	-- ProjLOAD - Solaris 10 per Project Load Statistic(s)	-[Ok]-
T3stat	-- T3 Storage Statistics (Warning: need configuration on STAT-service side!)	-[Ok]-
TaskLOAD	-- TaskLOAD - Solaris 10 per Task Load Statistic(s)	-[Ok]-

[http://localhost/cgi-bin/WebX.mySQL/dim_STAT/x.new](#)
100%



Any stats become Add-On?...

- Any! :-)
- Single-Line
 - > like “vmstat”
- Multi-Line
 - > like “iostat”
- Regular output!
 - > output is flushed per time interval
 - > fixed number of columns
 - > well defined data separators and headers (if any)



Example: ZFS Memory Usage Add-On

- before update4:
 - > # echo "arc::print -d c_max" | mdb -k <== Max limit
 - > # echo "arc::print -d size" | mdb -k <== Current
- after update4:
 - > # kstat -m zfs -s c_max -p
 - > # kstat -m zfs -s size -p
- # ZFS_mem.sh 10
 - c_max: 273212833 size: 136606416
 - c_max: 273212833 size: 136606446
 - c_max: 273212833 size: 136606006
 - ...

Example Add-On continue...

1.

Database: [RSBank]

[\[Home \]](#) [\[Preferences \]](#) [\[LOG Admin \]](#)

dim_STAT Integrate New Add-On STAT

STAT Name	<input type="text" value="ZFSstat"/> (One Word)
Columns per Line	<input type="text" value="2"/>
Type of output	<input type="text" value="Single-Line"/> per measurement

Example Add-On continue...

Database: [RSBank]

1.

2.

Database: [RSBank]

dim_STAT Integrate Ne

Database: [RSBank]

[\[Home \]](#) [\[Preferences \]](#) [\[LOG Admin \]](#)

dim_STAT Integrate New Add-On STAT

STAT Name	<input type="text" value="ZFSstat"/> (One Word)
Columns per Line	<input type="text" value="2"/>
Type of output	Single-Line per measurement

Add-On STAT Name	ZFSstat
Description	<input type="text" value="ZFSstat Statistic(s)"/>
Shell Command	<input type="text" value="ZFSstat %i"/> %h - hostname, %i - interval(sec.), %p - parameter(s)
Ignore Line(s)	Any line from pattern(s): <input type="text"/>

DB ColumnName	DataType	Column# in input Line	Short Name	Full Name	Use in Multi-Host Analyze
id	int	-	ID	Stat ID	-
sno	int	-	Sno	#Serie STAT	-
<input type="text" value="c_max"/>	Integer	<input type="text" value="2"/>	<input type="text" value="MemLimit"/>	<input type="text" value="ZFS Memory Limit"/>	No
<input type="text" value="c_size"/>	Integer	<input type="text" value="4"/>	<input type="text" value="MemUsage"/>	<input type="text" value="ZFS Memory Usage"/>	Yes

Example Add-On continue...

Database: [RSBank]

1.

2.

Database: [RSBank]

dim_STAT Integrate Ne

[Home] [Preferences] [LOG Admin]

dim_STAT Integrate New Add-On STAT

STAT Name (One Word)

Columns per Line

Type of output Single-Line per measurement

New Add-ON STAT **Reset**

Add-On STAT Name ZFSstat

Description

Shell Command %h - hostname, %i - interval(sec.), %p - parameter(s)

Ignore Line(s) Any line from pattern(s):

DB ColumnName	DataType	Column# in input Line	Short Name	Full Name	Use in Multi-Host Analyze
id	int	-	ID	Stat ID	-
sno	int	-	Sno	#Serie STAT	-
<input type="text" value="c_max"/>	Integer	<input type="text" value="2"/>	MemLimit	ZFS Memory Limit	No
<input type="text" value="c_size"/>	Integer	<input type="text" value="4"/>	MemUsage	ZFS Memory Usage	Yes

Create **Reset**

3.

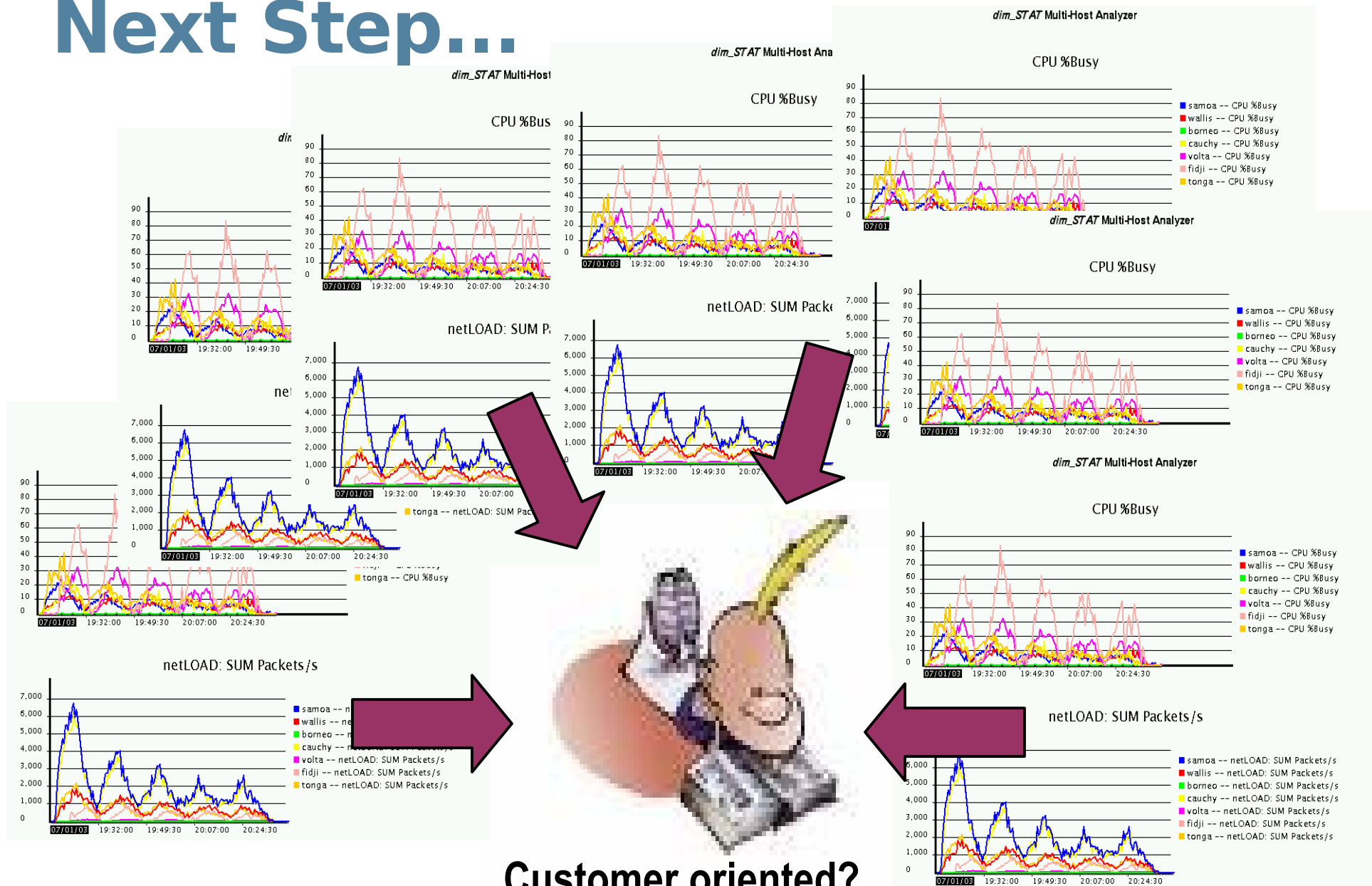
/etc/STATsrv/access

command ZFSstat /usr/bin/ZFS_mem.sh

Work Cycle Demo

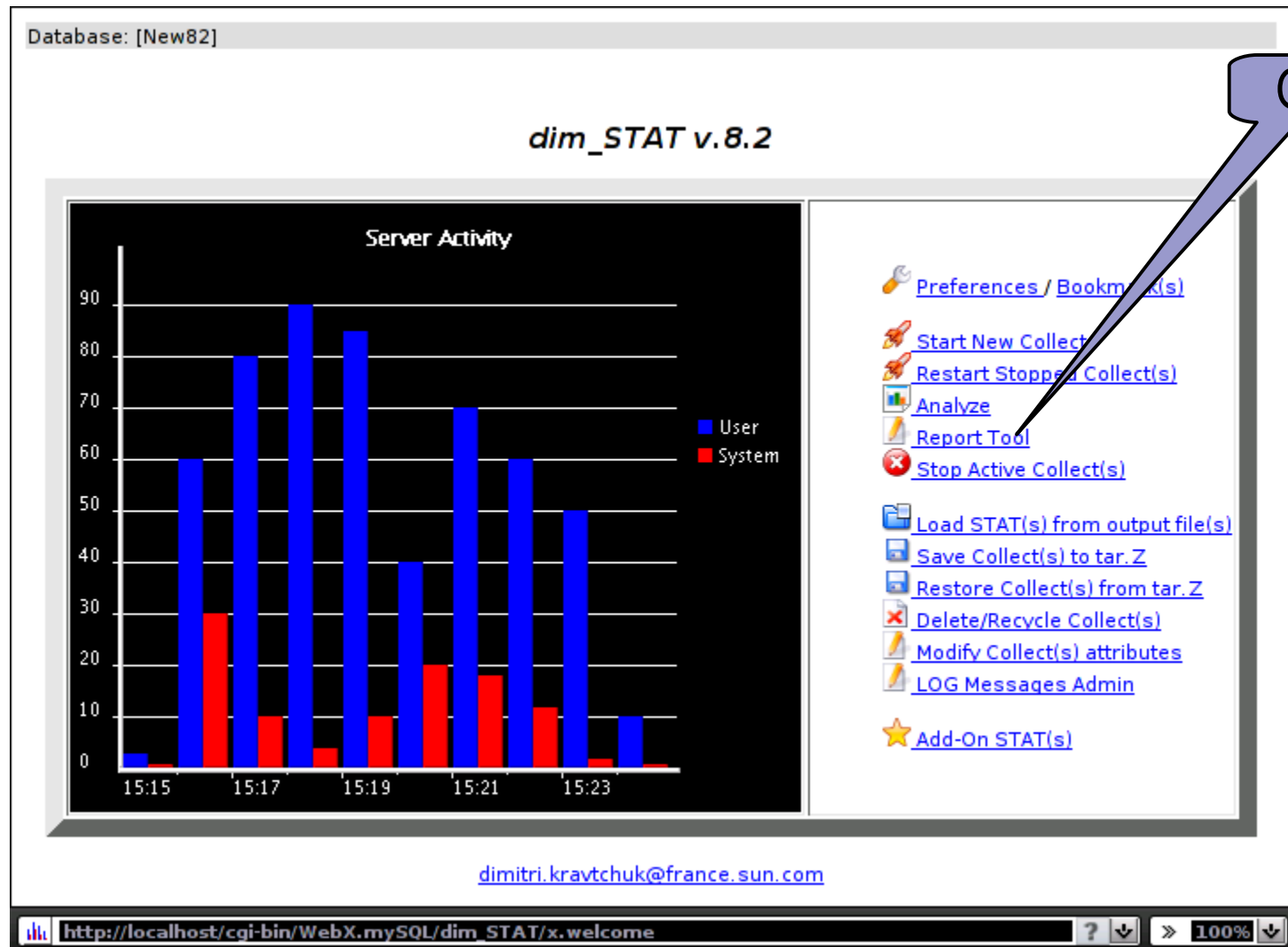
- Preferences
- Start Collect
- Analyze
- Bookmarks
- Add-Ons
- Log messages
- etc...

Next Step...



Customer oriented?

Report Tool

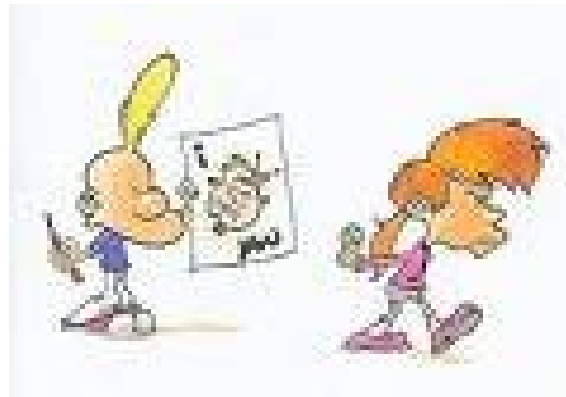


Go!



Reporting needs...

- Several versions of the same story:
 - > Confidentiality level (Customer, Partner, Internal, etc.)
 - > Language (English, French, German, etc.)
- Copy & Paste graphs is too painful...
- Automate whatever possible...
- Simplified formatting, teamwork and publishing...
- etc...

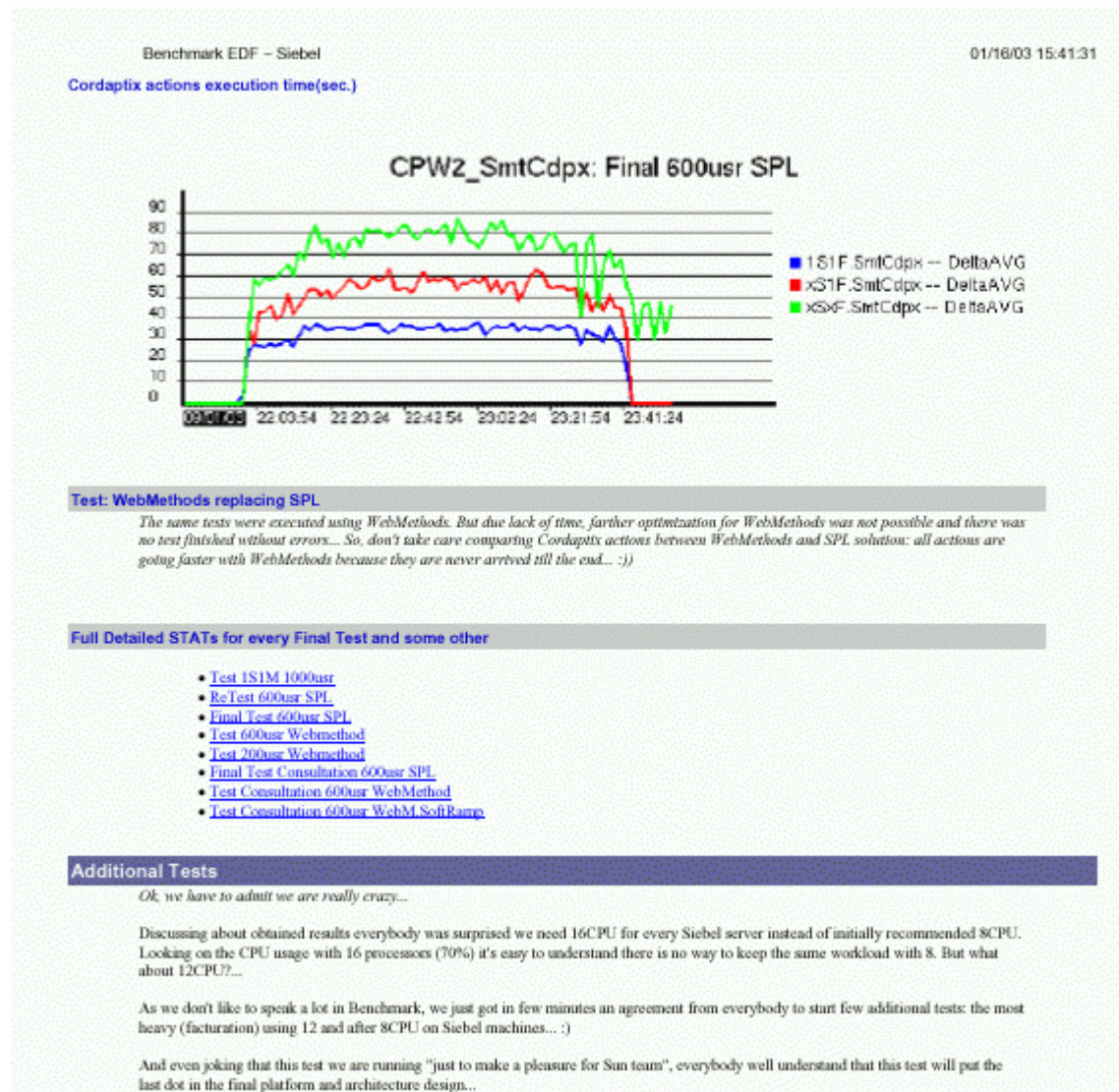


Report Tool

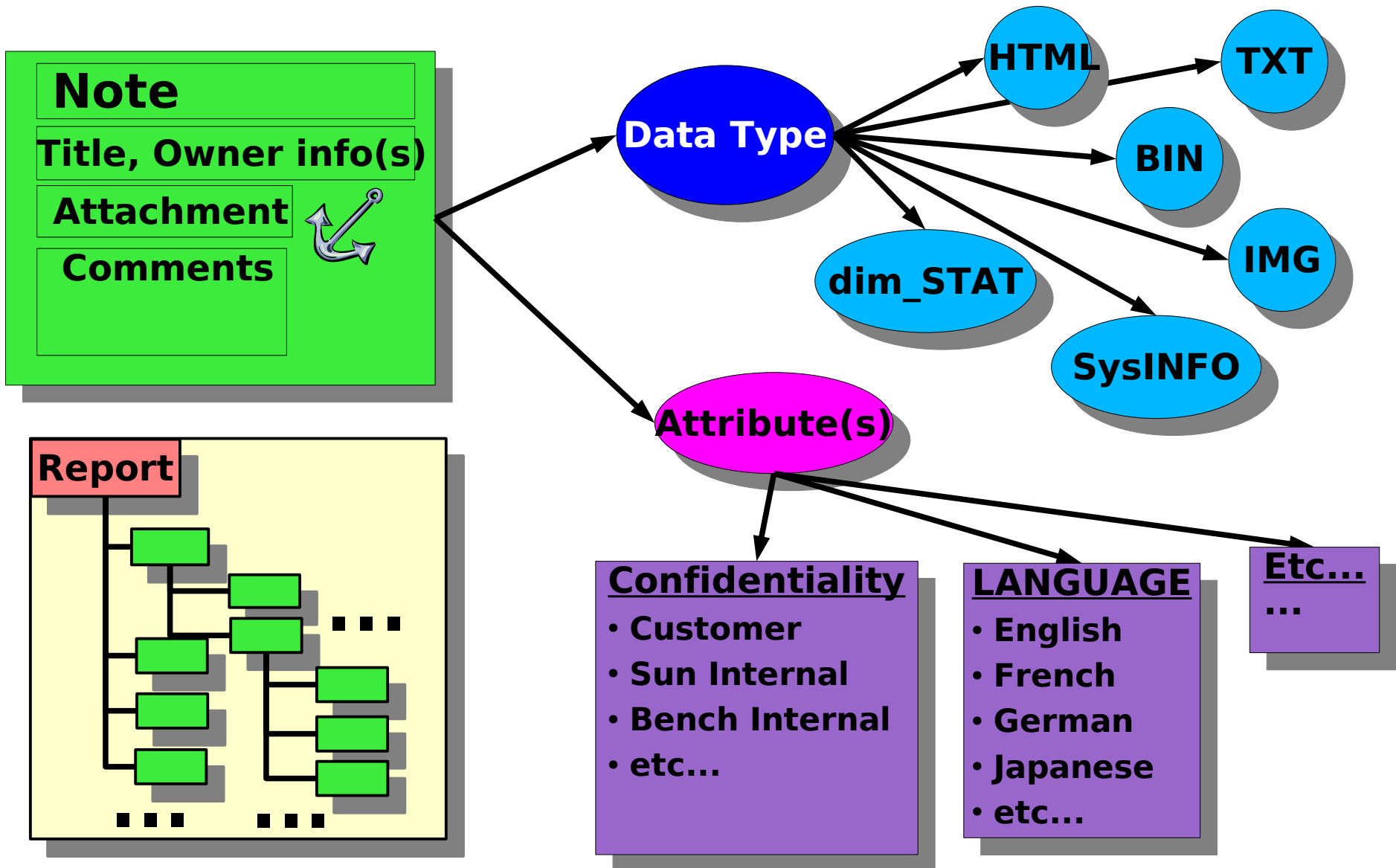
- Web Interface, Database driven
- Report Import / Export
- **Mostly Automated** Processing
- Supported data types:
 - > dim_STAT Collect
 - > SysINFO
 - > Text, HTML
 - > Image, Binary
 - > HTAR (tar archive with HTML documents)
- **Several Views** of the same report
- HTML and/or PDF output



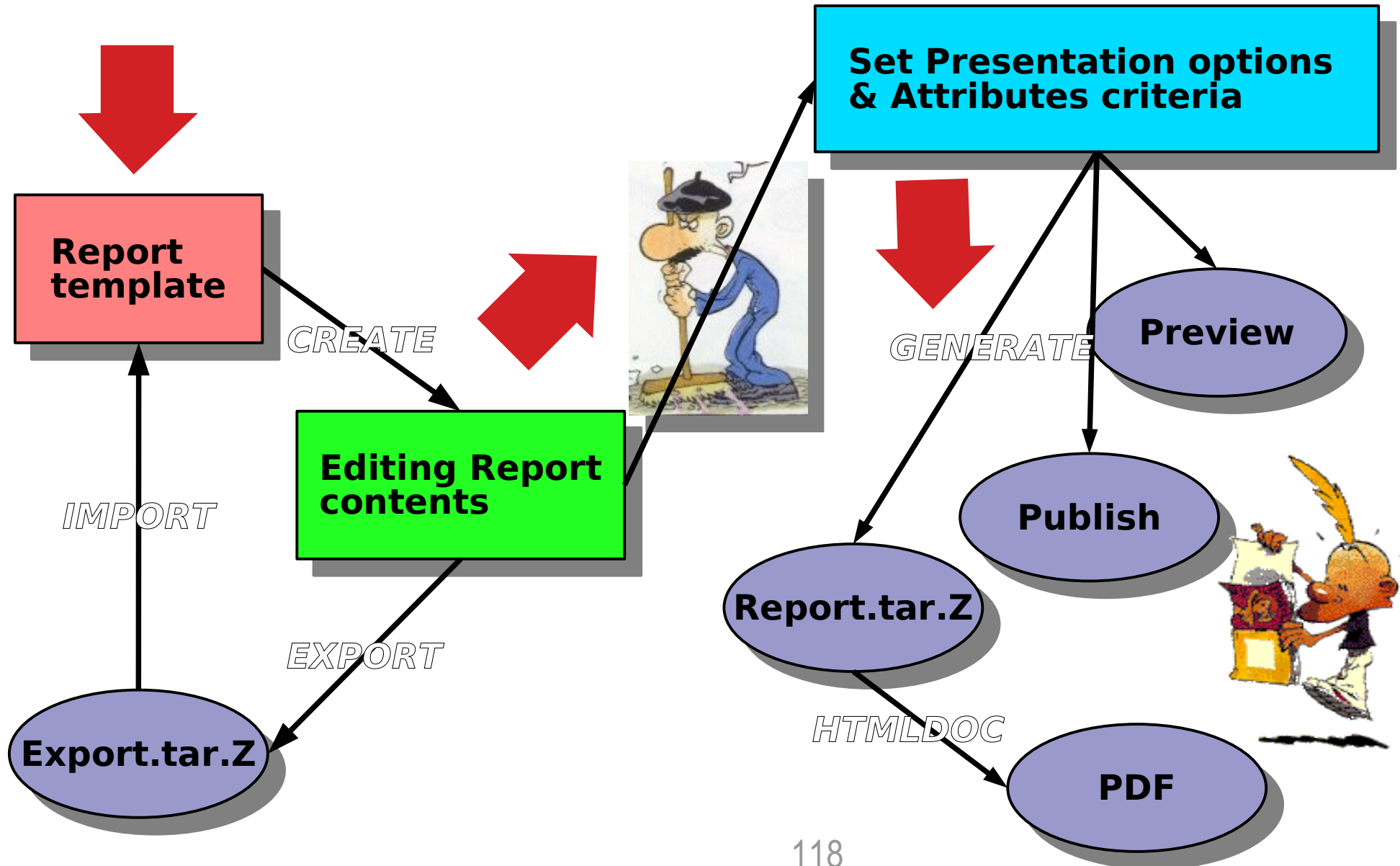
Example of Report page...



Report Tool architecture...



Report Working cycle



Example: New Report

Start a New Report

Report ID:	<input type="text" value="20071002"/>
Title:	<input type="text" value="Test Report"/>
Owner:	<input type="text" value="Dimitri"/>
Chart Comments:	<input type="text" value="SSC Team, 2007"/> <input type="text" value="Sun Microsystems Inc."/>
Use: <input type="text" value="WCB Template"/>	

???

Test Report

Benchmark Information

Customer Name(s):

NDA:

Contact Information:

Dates:

Keywords:

Hardware Configuration

Server(s):

Storage:

Software Configuration

System:









Application(s):

Benchmark



Example: New Note

New Note

-  [Text](#)
 -  [Image](#)
 -  [HTML](#)
 -  [Binary File](#)
 -  [dim_STAT-Collect](#)
 -  [dim_STAT-Snapshot](#)
 -  [SysINFO](#)
 -  [HTML.tar.Z Archive](#)
-



Report: dim_STAT Collect

- **No** copy & paste!... :-)
- Just open new **dim_STAT Note**:
 - > choose server + DB name
 - > choose STATs/ Bookmarks
 - > choose time periods
 - > **GO!**



- Take your coffee and leave the machine working for you (finally)...

Report screen-shot...

2006-12-13 17:55 [Connq2 Test 4xref 2xPromo jvm (400+750s)]

Thu, 14/Dec

Main events:

- Oracle: customer disco
- restore previous init. or
- restart tests
- Test 400ref+70promo:
- Test 400ref alone: ok!
- Test 100promo alone: c
- Sun appserver: continu

SysINFO: neel @

Warning: RAM Use

```
neel # mdb -k
> ::memstat
Page Summary
.....
Kernel
Anon
Exec and libs
Page cache
Free (cachelist)
Free (freelist)

Total
Physical
>
```

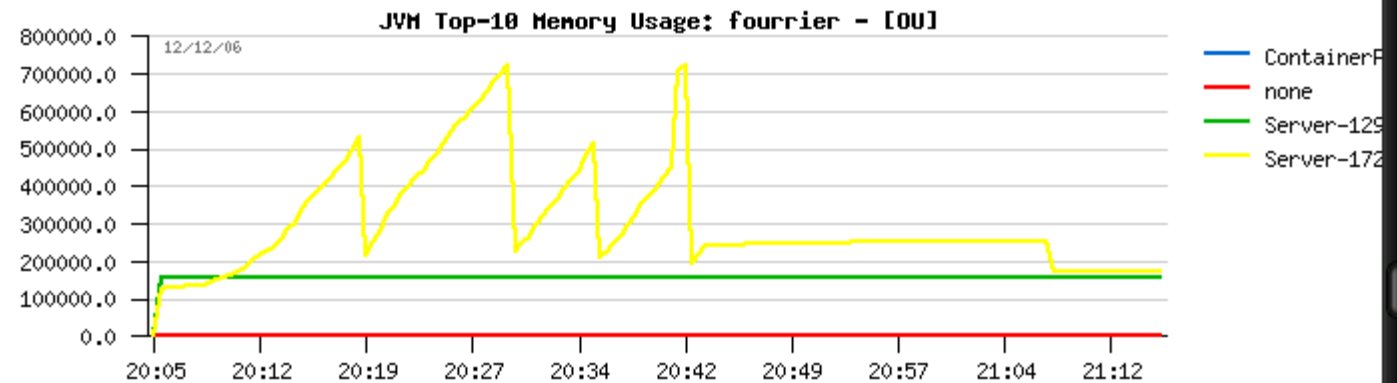
57 GBytes u

SysINFO: fourrier @

STATs High load tests, c

- 2006-12-14
- 2006-12-14
- 2006-12-14

JVM Top-10 Memory Usage



Network Bytes/sec



http://localhost/cgi-bin/WebX.mysql

http://localhost/WCB_Reports/Report_1214094/3704_dim_STAT_6.html

100%

Demo: Full Work Cycle

- Create New Report
- Insert Text, Image, SysINFO, BIN
- Insert dim_STAT Collect graphs
 - > based on log messages
 - > based on time cycles
- Move/Delete/Re-Edit Notes
- Preview
- Generate
- Export/Import

Pending features...

- “Grid Ready” (hundreds of hosts)
- Customized Alerts
- Health Checker
- Automated Perf.Analyzer
- Dash board
- etc...

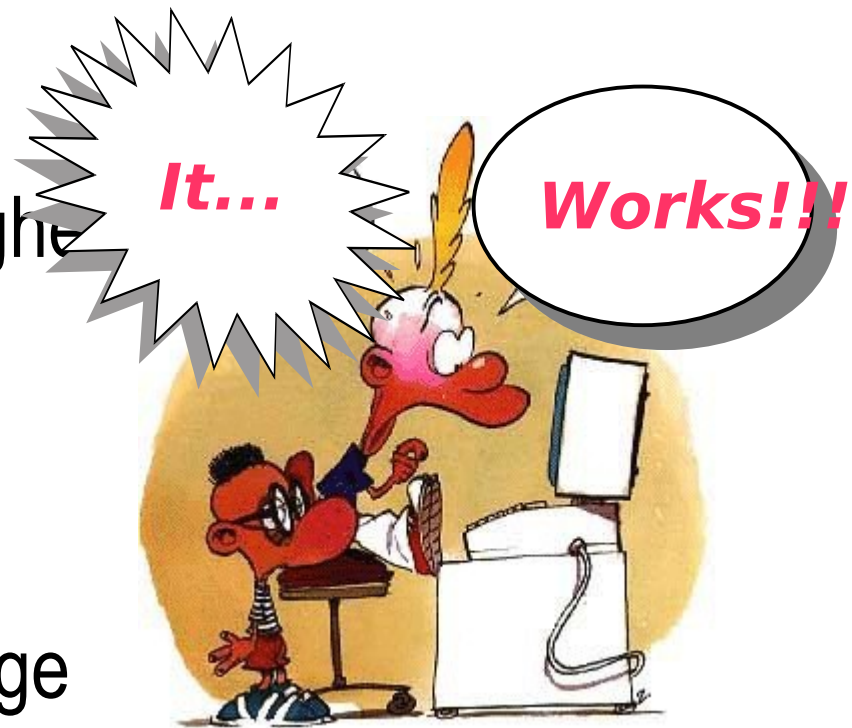


Key Notes

- Freeware/GPL!
- Solaris/SPARC (since 2.6 and higher)
- Solaris 10 x86, Linux/x86
- 5 min. install & run
 - > all software is pre-bundled!
- 50MB disk space, 0.1% CPU usage

Download:

- SWAN: <http://goldgate.france>
- Internet: <http://dimitrik.free.fr>



Some Facts

- 10+ Years! :-)
- Sun: ToolsCD, all BMK Centers, etc.
- Alcatel (World-wide and pre-installed by CRS)
- Orange, EDF, Renault, Nokia, Motorola, etc...
- Ready for Zones / LDOMs :-)
- Downloads since Jul.2007...



“your program is one of Sun's best kept secrets”
Tom Alling, tecsol.com



Contact

- Dimitri KRAVTCHUK
dimitri@sun.com
- Matthieu BORDONNE
matthieu.bordonne@sun.com

Sun Solution Center
Paris, FRANCE



N.B. *A Tool will never replace an Engineer!!!*