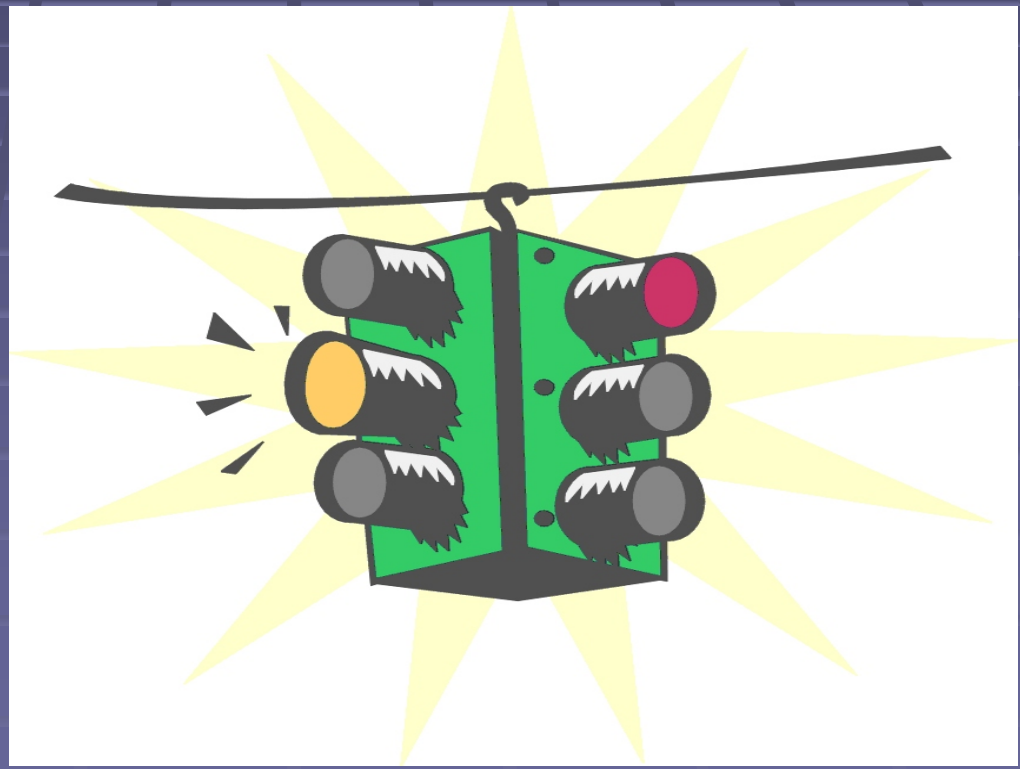
Do-It-Yourself Session Monitoring with OTop

How to use a custom top-like program to monitor Oracle

Waiting, Waiting, Waiting

- While driving - when you hit traffic

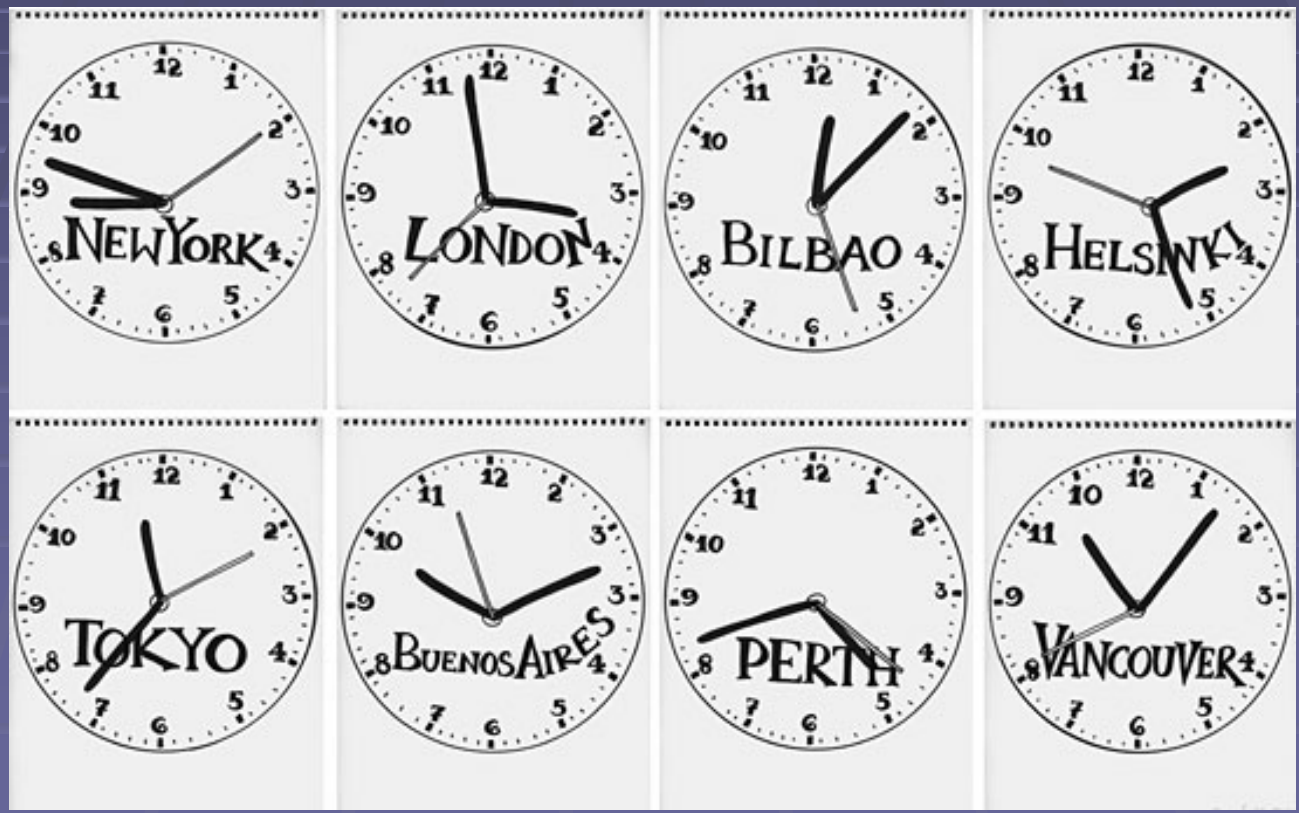


Waiting, Waiting, Waiting

- When you're on line at the bank

Waiting, Waiting, Waiting

- When you're coordinating a meeting in different timezones



Is Waiting Important?

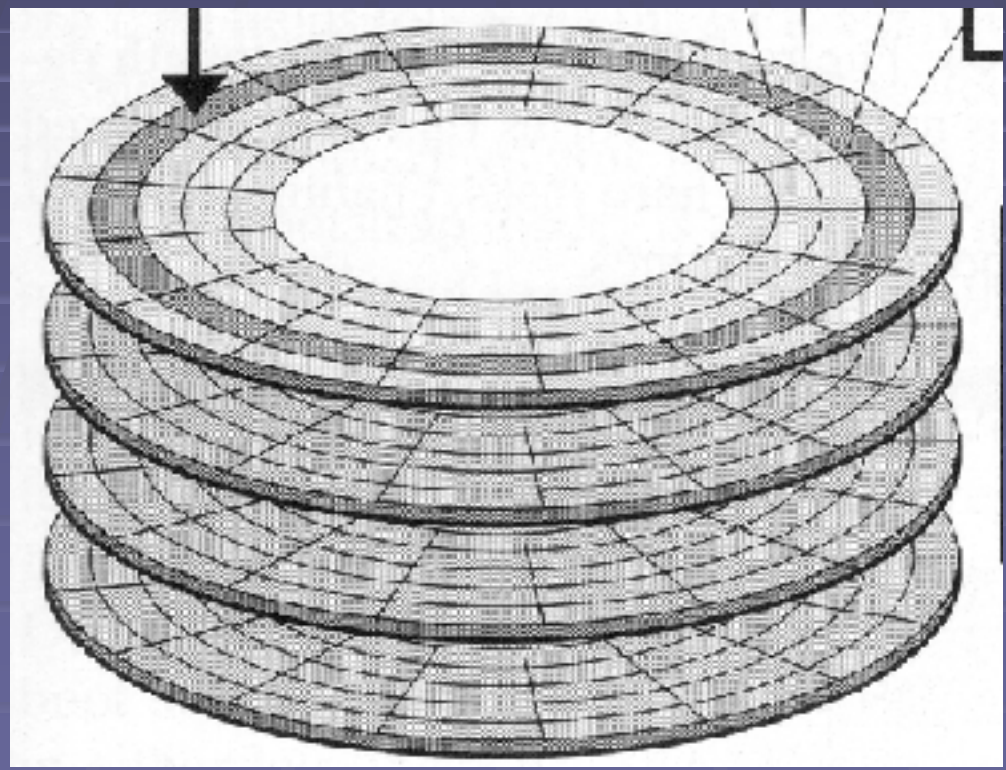
- When you're at the doctor's office
- Whenever you have things planned back to back
- Cascading lateness
- Little wastes of time between events
- Timeliness requires organization, to use up all those little pieces of time

Waiting in Computing

- Unix process context switching
  - Waiting for other processes
  - Time to switch from one to another
  - System/kernel time not available to user procs

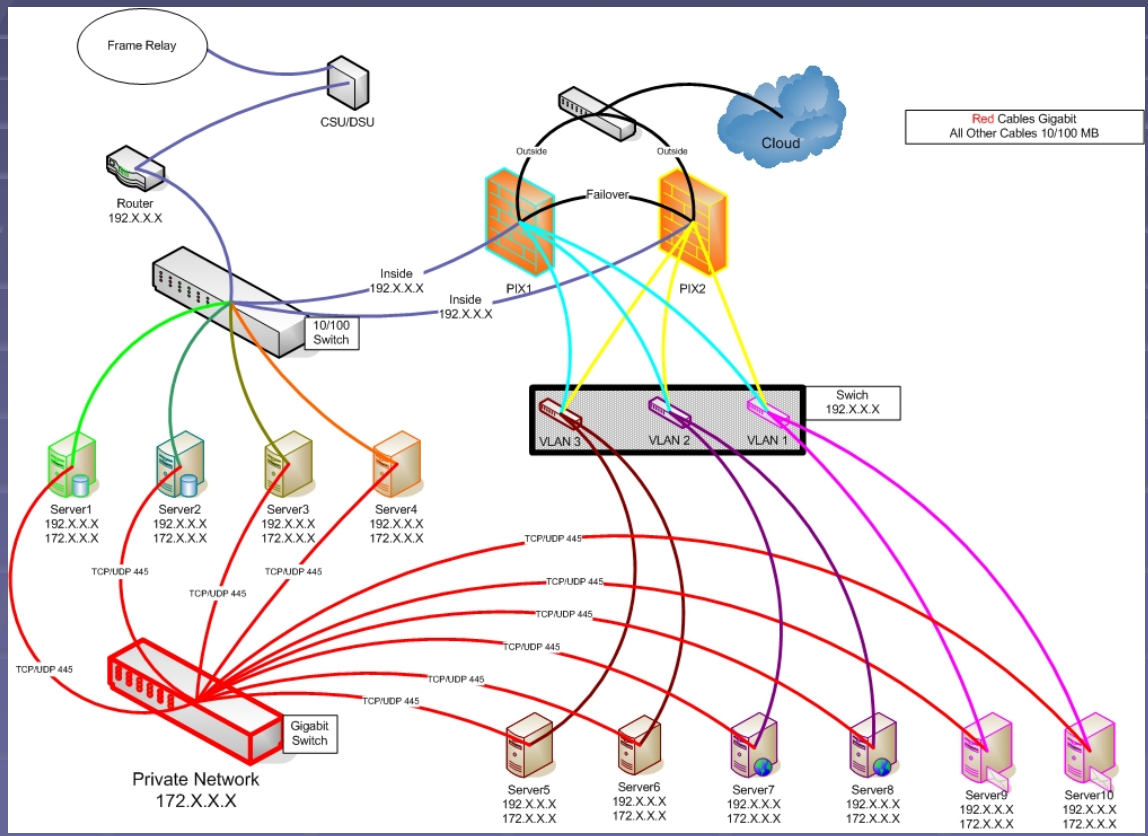
Waiting in Computing

- Disk I/O Subsystem



Waiting in Computing

- Network Latency



Powered by Heavyweight Internet Group

Computing : Dining

- Disk I/O : Seating Capacity
  - Waiting to get a table can be frustrating
- CPUs : Cooks
  - Waiting for your waitor, if they don't manage their tables well, slow!
- Memory : Kitchen size
  - You need enough cooks to handle what
- Processes/Threads : Waiters
  - Need to work in parallel, minmize context switches!

Oracle Waiting

- Organized into named events
- Exposed via Data Dictionary views

Oracle Wait Events: Disk I/O

- db file sequential read
  - Single block read (undo, rlbk, indx)
  - Table access by rowid
- db file scattered read
  - Blocks read & “scattered” as written to buffer cache

Oracle Wait Events: Sorting

- direct path read
  - Also during parallel query + hash joins
- direct path write
  - Create table as select
  - Parallel dml

Oracle Wait Events: SGA

- buffer busy waits (read by other session)
  - another session is accessing that block
- free buffer waits
  - No free buffers in buffer cache
- library cache pin
  - Compiling/parsing PL/SQL + Views

Oracle Wait Events: Locks

- latch free
  - Serialize access, but no waiting queue
  - Process can spin, and request latch again
- Enqueue
  - Semaphore, serialize access to something

Oracle Wait Events: Other

- db file parallel read
- db file parallel write
- control file parallel write
- library cache lock
- Log buffer space
- log file parallel write
- log file sequential read
- log file switch

Oracle Wait Events: Other

- log file switch completion
- log file sync
- SQL*Net message from client
- SQL*Net message to client
- RAC events

How Do We Watch Events?

- ENTER: Oracle Wait Interface!
- Oracle exposes db kernel memory structures via data dictionary views
- Allows us to use SQL to keep an eye on things
- Non-transactional, constantly changing
  - Even while we are querying!!

OWI: Important Views

- V$event_name
- V$system_event
- V$session_event
- V$session_wait
- V$session_wait_history (10g)
- Others?

V$event_name

```
SQL> desc v$event_name;
 Name                      Null?    Type
------------------------- -------- -----------
EVENT#                             NUMBER
EVENT_ID                           NUMBER
NAME                               VARCHAR2(64)
PARAMETER1                         VARCHAR2(64)
PARAMETER2                         VARCHAR2(64)
PARAMETER3                         VARCHAR2(64)
WAIT_CLASS_ID                      NUMBER
WAIT_CLASS#                        NUMBER
WAIT_CLASS                         VARCHAR2(64)
```

V$event_name example

```
SQL> SELECT event#, name
  2  FROM v$event_name WHERE name like 'db%';

    EVENT# NAME
---------- ------------------------------
         6 dbms_file_transfer I/O
       115 db file sequential read
       116 db file scattered read
       117 db file single write
       118 db file parallel write
       119 db file parallel read

6 rows selected.
```

V$system_event

```
SQL> desc v$system_event;
 Name                      Null?    Type
 ------------------------- -------- --------
 EVENT                              VARCHAR2(64)
 TOTAL_WAITS                        NUMBER
 TOTAL_TIMEOUTS                     NUMBER
 TIME_WAITED                        NUMBER
 AVERAGE_WAIT                       NUMBER
 TIME_WAITED_MICRO                  NUMBER
 EVENT_ID                           NUMBER
 WAIT_CLASS_ID                      NUMBER
 WAIT_CLASS#                        NUMBER
 WAIT_CLASS                         VARCHAR2(64)
```

V$system_event example

```
SQL> SELECT event, average_wait
  2  FROM v$system_event
  3  WHERE event like 'db file%';

EVENT                                    AVERAGE_WAIT
---------------------------------------- ------------
db file sequential read                           .67
db file scattered read                           1.12
db file single write                              .01
db file parallel write                            .44
db file parallel read                            1.52
```

V$session_event

```
SQL> desc v$session_event;
 Name                      Null?    Type
 ----------------- -------- -------
 SID                                NUMBER
 EVENT                              VARCHAR2(64)
 TOTAL_WAITS                        NUMBER
 TOTAL_TIMEOUTS                     NUMBER
 TIME_WAITED                        NUMBER
 AVERAGE_WAIT                       NUMBER
 MAX_WAIT                           NUMBER
 TIME_WAITED_MICRO                  NUMBER
 EVENT_ID                           NUMBER
 WAIT_CLASS_ID                      NUMBER
 WAIT_CLASS#                        NUMBER
 WAIT_CLASS                         VARCHAR2(64)
```

V$session_event example

```
SQL>  SELECT sid, event, average_wait FROM v$session_event
  2  WHERE event like 'db file%';

  SID EVENT                          AVERAGE_WAIT
----- ------------------------------ ------------
  137 db file sequential read                2.42
  143 db file sequential read                3.03
  154 db file sequential read                1.64
  154 db file scattered read                  .93
  155 db file sequential read                2.26
  156 db file sequential read                 .16
  157 db file sequential read                 .15
  157 db file scattered read                 1.04
  157 db file parallel read                   .01
  160 db file parallel write                  .44

10 rows selected.
```

V$session_wait

```
SQL> desc v$session_wait
 Name                      Null?    Type
 ----------------------- -------- ---------
 SID                                NUMBER
 SEQ#                               NUMBER
 EVENT                              VARCHAR2(64)
 P1TEXT                             VARCHAR2(64)
 P1                                 NUMBER
 P1RAW                              RAW(4)
 P2TEXT                             VARCHAR2(64)
 P2                                 NUMBER
 P2RAW                              RAW(4)
 P3TEXT                             VARCHAR2(64)
 P3                                 NUMBER
 P3RAW                              RAW(4)
 WAIT_CLASS_ID                      NUMBER
 WAIT_CLASS#                        NUMBER
 WAIT_CLASS                         VARCHAR2(64)
 WAIT_TIME                          NUMBER
 SECONDS_IN_WAIT                    NUMBER
 STATE                              VARCHAR2(19)
```

V$session_wait example

```
SQL> insert into sean_test select * from sean_test;

SQL> SELECT event
  2 FROM v$session_wait
  3 WHERE sid = ‘140’;

EVENT
----------------------------

free buffer waits
```

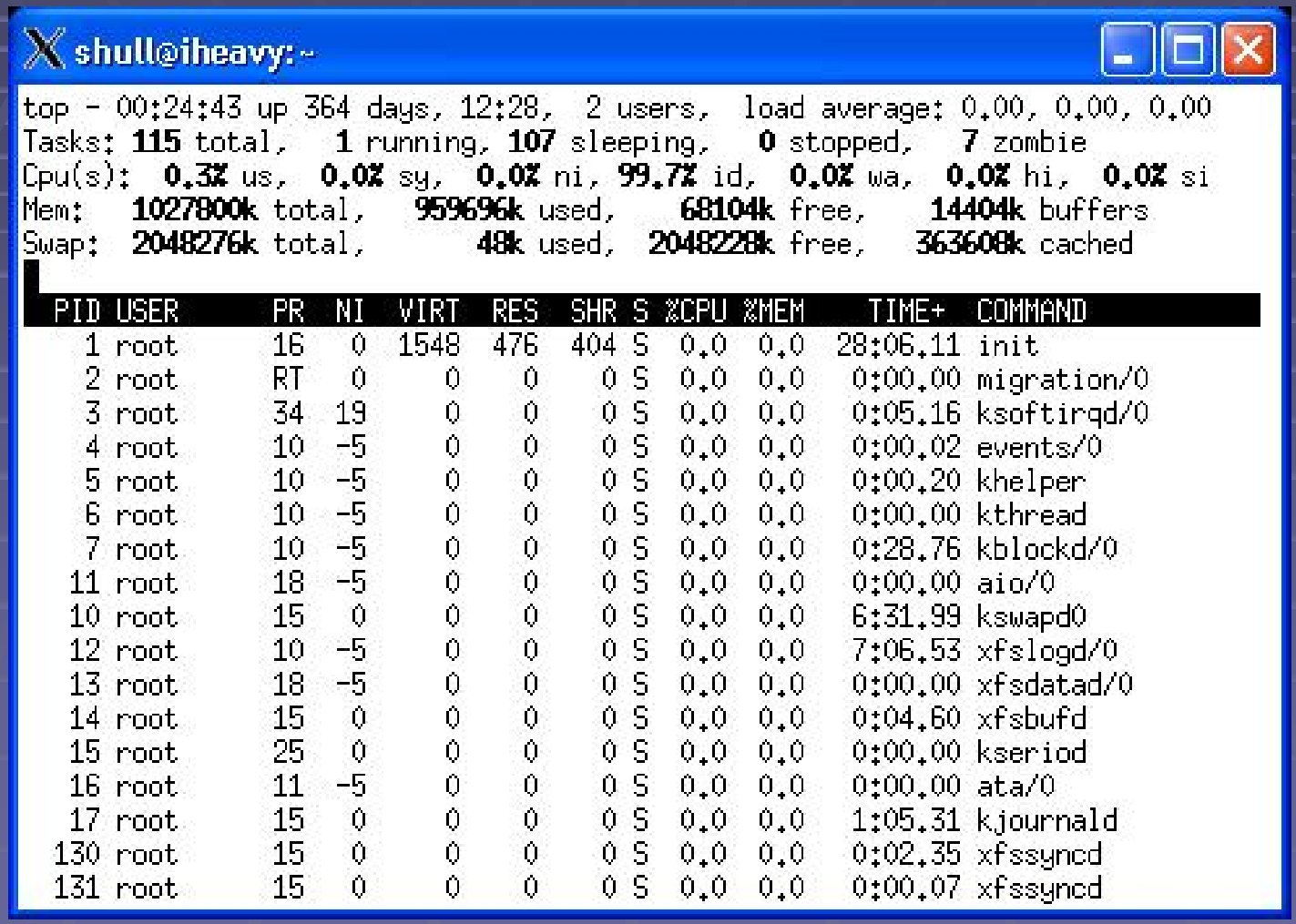
Manual Querying: Problems

- Constantly changing
- Don’t have historical data
  - 10g has V$session_wait_history
- Other?

Is There A Better Way?

- Want a polling tool which is low-impact
- Want to use via SSH login
- Easy install, Curses library might be ideal
- Do we have examples to model something new?

We All Use + Love TOP!



Why We Like Top

- At a glance view of heaviest processes
- Memory Usage
- Disk Usage
- Load Average
- CPU Busy/Idle
- Add/Remove/Sort columns
- Control the update interval
- Color highlighting

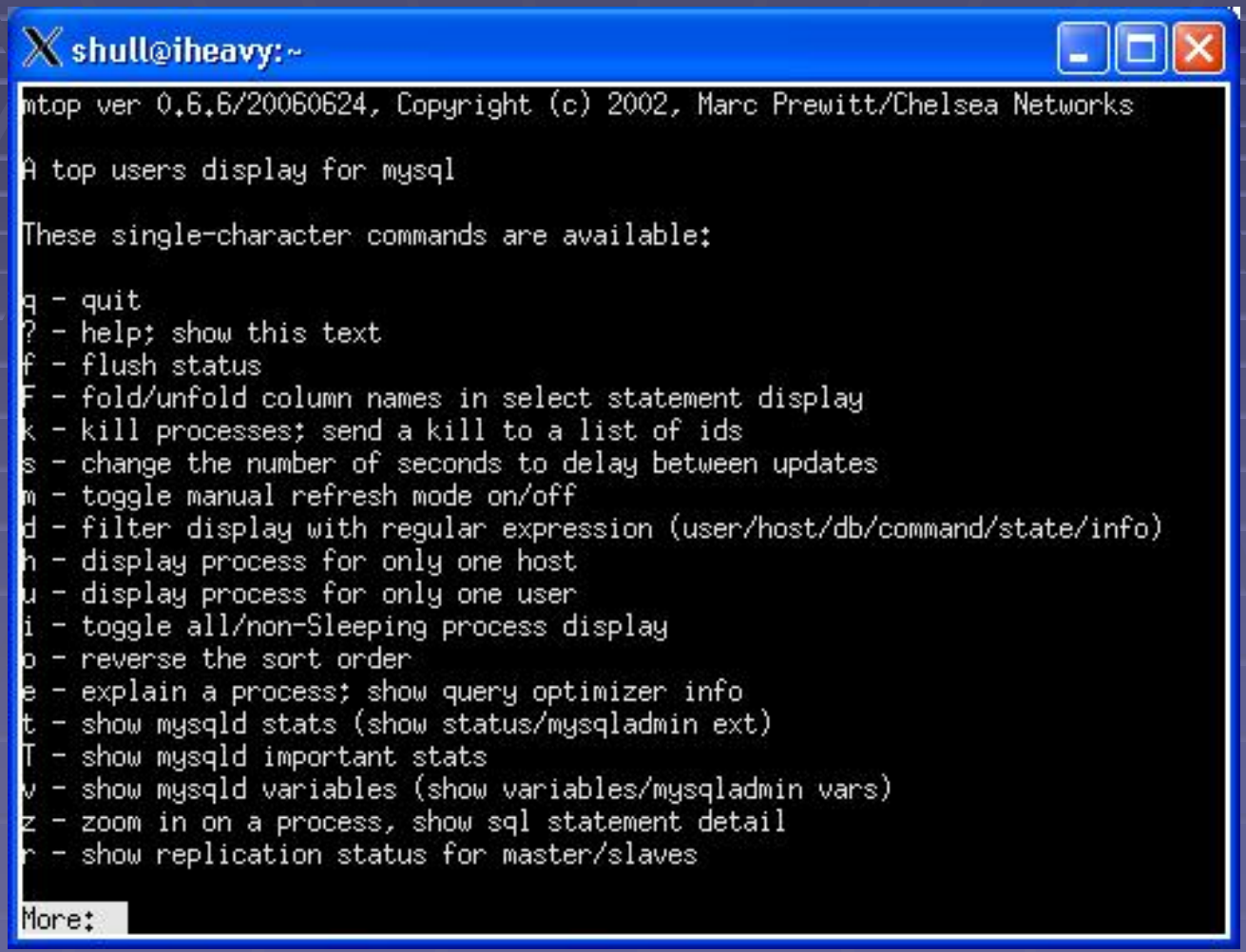
Have You Heard of mtop?

shull@iheavy:~

```
load average: 0.00, 0.01, 0.00 mysqld 4.1.13-standard up 29 day(s), 13:37 hrs
15 threads: 1 running, 0 cached. Queries/slow: 3.3M/0 Cache Hit: 99.02%
Opened tables: 58.5K  RRN: 63.6M  TLW: 38  SFJ: 909  SMP: 0  QPS: 0

ID       USER     HOST            DB          TIME   COMMAND STATE        INFO
9984     horde    localhost       horde       3548   Sleep
12284    horde    localhost       horde       2948   Sleep
9977     horde    localhost       horde       2006   Sleep
9985     horde    localhost       horde       1748   Sleep
12278    horde    localhost       horde       1706   Sleep
7884     horde    localhost       horde       1448   Sleep
9970     horde    localhost       horde       1148   Sleep
12280    horde    localhost       horde       848    Sleep
12281    horde    localhost       horde       548    Sleep
12286    horde    localhost       horde       248    Sleep
56167    sugar    localhost       sugar       203    Sleep
56168    sugar    localhost       sugar       55     Sleep
56187    root     localhost       nyougsugar  22     Sleep
56189    root     localhost       nyougsugar  22     Sleep
56165    mysqltop localhost                           Query                show
---
```

mtop – help screen



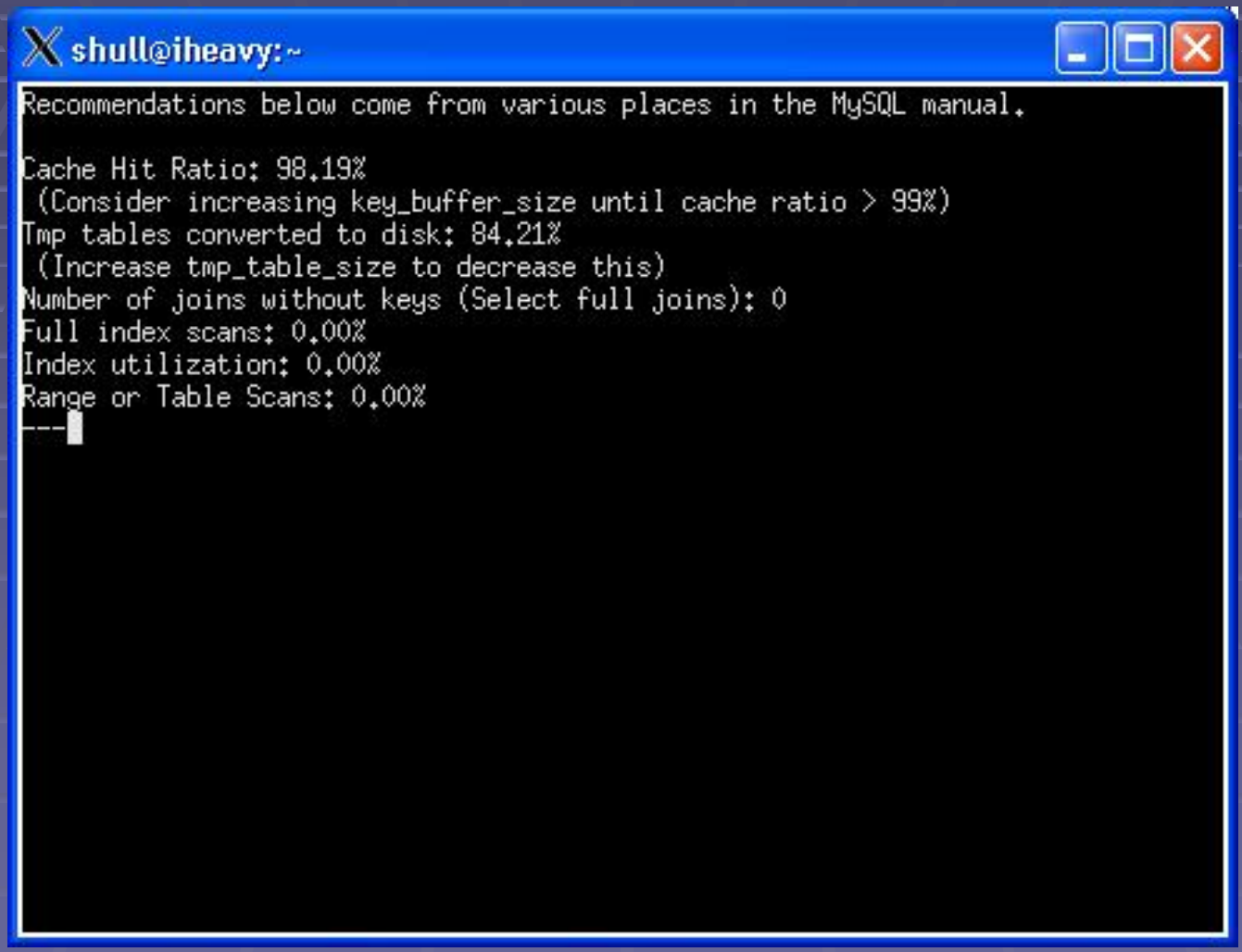
mtop - variables

shull@iheavy:~

back_log:	50	log_slave_updates:	OFF
basedir:	/	log_slow_queries:	OFF
binlog_cache_size:	32,768	log_update:	OFF
bulk_insert_buffer_size:	8,388,608	log_warnings:	1
character_set_client:	latin1	long_query_time:	10
character_set_connection:	latin1	low_priority_updates:	OFF
character_set_database:	latin1	lower_case_file_system:	OFF
character_set_results:	latin1	lower_case_table_names:	0
character_set_server:	latin1	max_allowed_packet:	1,048,576
character_set_system:	utf8	max_binlog_cache_size:	4,294,967,295
character_sets_dir:	/usr/share/mysql/charsets/	max_binlog_size:	1,073,7
collation_connection:	latin1_swedish_ci	max_connect_errors:	10
collation_database:	latin1_swedish_ci	max_connections:	100
collation_server:	latin1_swedish_ci	max_delayed_threads:	20
concurrent_insert:	ON	max_error_count:	64
connect_timeout:	5	max_heap_table_size:	16,777,216
datadir:	/var/lib/mysql/	max_insert_delayed_threads:	20
date_format:	%Y-%m-%d	max_join_size:	4,294,967,295
datetime_format:	%Y-%m-%d %H:%i:%s	max_length_for_sort_data:	1,024
default_week_format:	0	max_relay_log_size:	0
delay_key_write:	ON	max_seeks_for_key:	4,294,967,295
delayed_insert_limit:	100	max_sort_length:	1,024
delayed_insert_timeout:	300	max_tmp_tables:	32
delayed_queue_size:	1,000	max_user_connections:	0
expire_logs_days:	0	max_write_lock_count:	4,294,967,295

More:

mtop - recommendations



Why mtop is popular

- At a glance view of processes
- Slow Query feedback
- db uptime
- Version, enabled features, add-ons
- Various Sorting options
- Locking, and related waiting activity
- Sorting activity, and related waiting
- OS information, load average etc

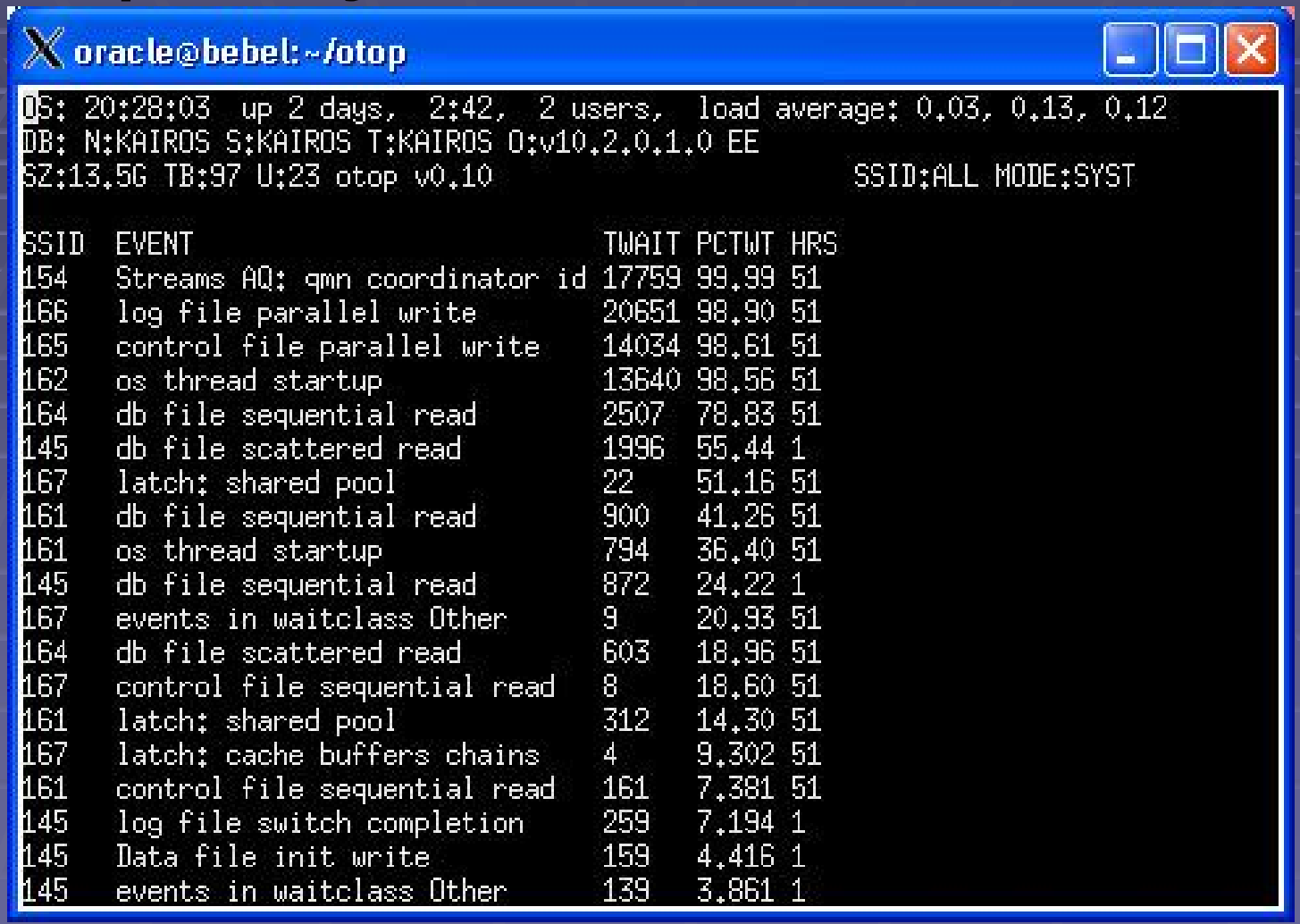
Why mtop is Popular

- Explain a process (SQL in a session)
- Kill a process (session)

Inspiration. We need otop!

- Session monitoring w/drill down
- System event monitoring w/drill down
- Various other info
  - Mem, disk, swap, load
  - SID, Hostname, etc

otop – system event monitoring

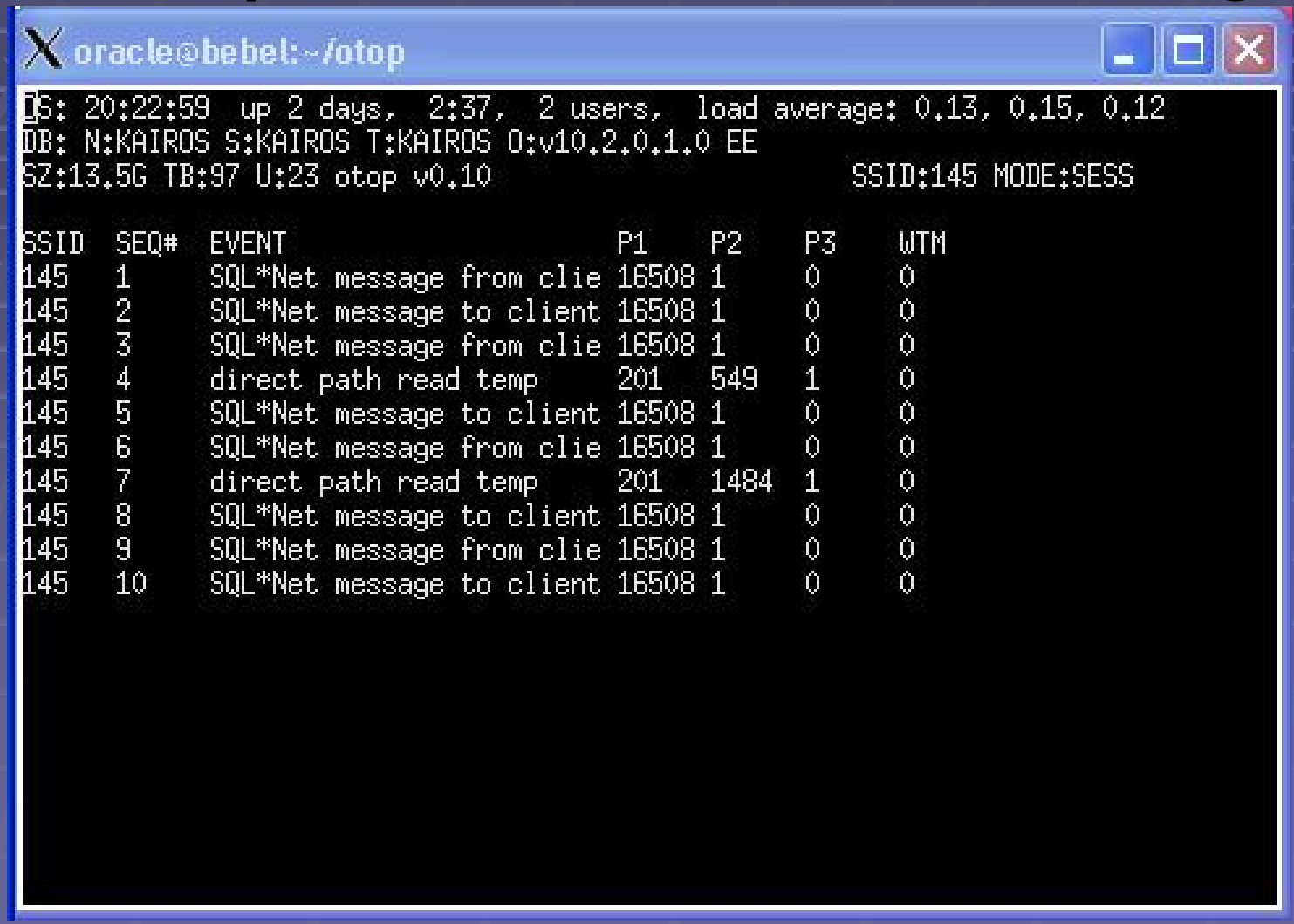


otop – system event monitoring

- otop was using this query:

```
SELECT a.sid,a.event,a.time_waited,
a.time_waited / c.sum_time_waited * 100 pct_wait_time,
round((sysdate - b.logon_time) * 24) hours_connected
FROM v$session_event a, v$session b,
 (select sid, sum(time_waited) sum_time_waited
  FROM v$session_event
  WHERE event not in (...)
  HAVING sum(time_waited) > 0 group by sid) c
WHERE a.sid = b.sid
AND   a.sid = c.sid AND   a.time_waited > 0
AND   (a.time_waited / c.sum_time_waited) < 1
ORDER BY pct_wait_time desc
```

otop – session monitoring



otop – session monitoring

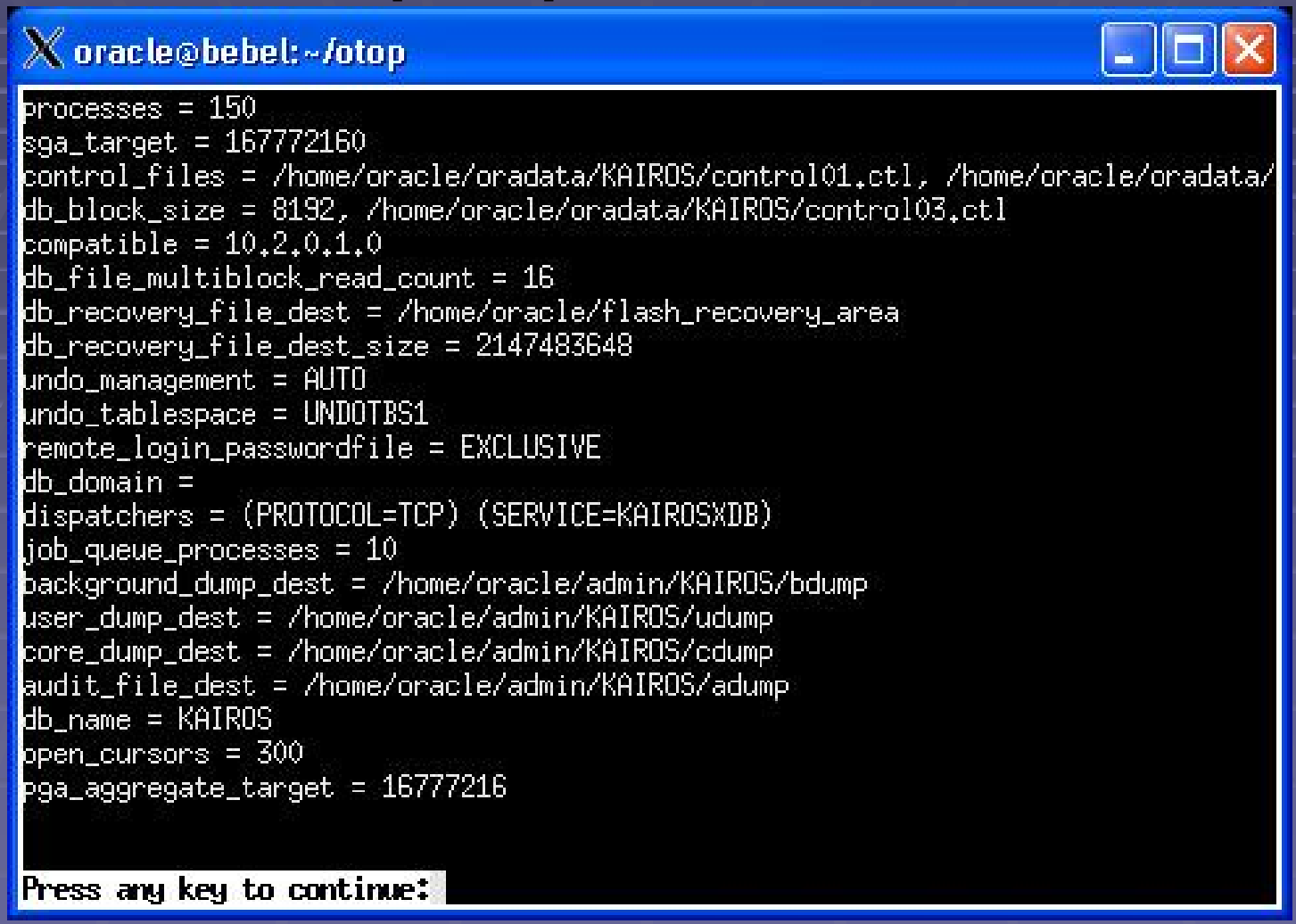
- Performing a sort
  - SELECT a, b, c FROM sean_test order by a;
- otop was monitoring using this query:
  - SELECT sid,seq#,event,p1,p2,p3, wait_time
    FROM v$session_wait_history
    WHERE sid = ?
    AND event not in (…)

otop - help

otop - help

- otop option “h”
- Show commands that interactively change otop’s behavior

otop - parameters



otop - parameters

- SELECT name, value
  FROM v$parameter
  WHERE isdefault = ‘FALSE’
- otop option “p”

otop - prerequisites

- Perl (tested on 5.8)
- Oracle (tested on 9.2,10.2)
- DBI (tested on 1.51)
- DBD::Oracle (tested on 1.17)
- Curses (tested on 1.14)

otop - installing

- Install Oracle 10g (Win32 or Unix)
- On Windows install ActiveState Perl
  - http://www.activestate.com/perl/
- perl –MCPAN –e shell
  - CPAN> install DBI
  - CPAN> install DBD::Oracle
  - CPAN> install Curses::Forms
- Download otop.tgz from here:
  - http://www.iheavy.com/otop/

otop - supported

- Linux 2.2, 2.4, 2.6 (any version w/Perl)
- Windows XP, 2k, NT (Activestate Perl)
- Oracle 9.2,10.2

otop - future

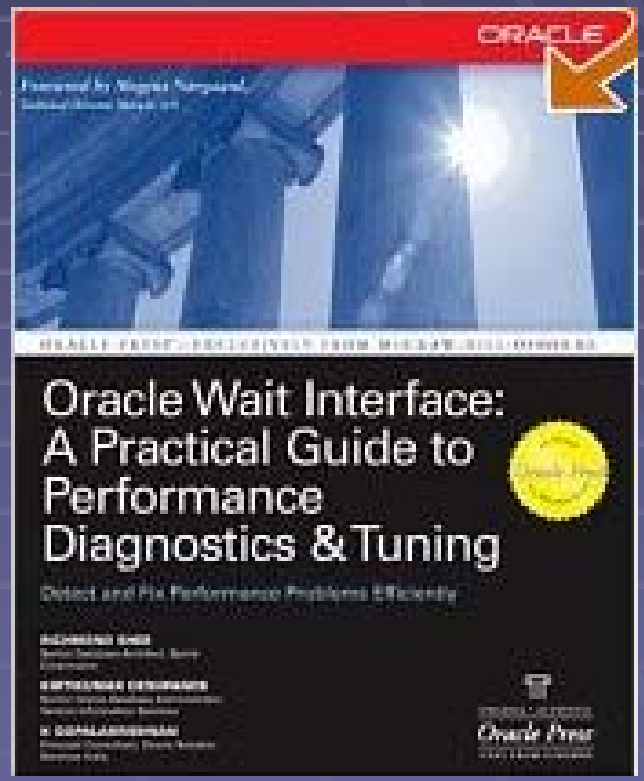
- How about toggling session history?
- How about toggling session trace?
- How about killing a session?
- How about thresholds for certain events?
- How about color coding for easier viewing at-a-glance?
- How about sorting
  - top disk I/O, top CPU, top MEM activity

otop - future

- How about db recommendations?
  - SGA settings, redolog file sizes, backups
  - Missing indexes, too many indexes
  - Missing bind variables in queries
  - Patch/upgrade recommendations
  - Security checks

OWI Bible

- Oracle Wait Interface
  - By Richmond Shee, Kirtikumar Deshpande and K. Gopalakrishnan



Rally the Troops!

Please Contact Me

- Interested in this project?
- Collaboration is a great way to learn about Oracle, and expand your skills.
- OS, Data Dictionary, Perl programming, and so on
- shull@iheavy.com
- 917.442.3939