NAME

saptune−note - Note definition files for saptune version **2**

DESCRIPTION

This man page documents the format of the Note definition files for saptune version **2**. If you still use version **1** of saptune please refer to the configuration files */etc/sysconfig/saptune-note-**

The saptune Note definitions will be searched in */usr/share/saptune/notes* for the saptune SAP Note definitions or */etc/saptune/override* for customer specific changes to the saptune SAP Note definitions or */etc/saptune/extra* for vendor or customer specific tuning definitions.
The **Note definition** files use the INI file format.
A comment line starts with #.
Lines starting with '[' indicate the begin of a new section.

SECTIONS

A section starts with a '[section_name]' keyword in the first line, followed by lines with options and comments.

The following section definitions are available and used in the saptune SAP Note definition files. Each of these sections can be used in a vendor or customer specific tuning definition placed in */etc/saptune/extra*.

List of supported sections: version, block, cpu, grub, limits, login, mem, pagecache, reminder, rpm, service, sysctl, vm

See detailed description below:

[version]

This section is a mandatory section and is used to display version, description and last change date of the underlying Note during saptune action 'list'.

Syntax:
 # <prefix>NOTE=<noteId> CATEGORY=<category> VERSION=<versionNo> DATE=<release date of used note and

Example:
VIP-NOTE=vip1 CATEGORY=VIP VERSION=5 DATE=16.04.2019 NAME="VIP: this is VIP Note 1, which contains Very Important Parameters"

All fields are separated by spaces. But please do not use spaces around the equal operator (=) of the fields. And please do not change the order of the fields.

The <noteId> must be a text string without spaces, which will be used as the unique identifier of this Note definition. It will be displayed during the action 'saptune note list' and used for all other actions, where the NoteID is needed as parameter.

The CATEGORY is for future use. So we do not have defined CATEGORIES at the moment. It must be a text string without spaces.

VERSION is a number that should indicate how many changes are done for this Note definition in the past. Only digits are allowed.

DATE is the date of the last changes.

NAME is the description of the Note, which will be displayed during the action 'saptune note list'
Attention: The note description from the field NAME must be placed in double quotes even if there are no spaces used inside the description.

[block]

The section "[block]" can contain the following options:

IO_SCHEDULER=*STRING*

The default I/O scheduler for single-queued block layer devices offers satisfactory performance for wide range of I/O task, however choosing an alternative scheduler may potentially yield better latency characteristics and throughput. "noop" is an alternative scheduler, in comparison to other schedulers it may offer more consistent performance, lower computation overhead, and potentially higher throughput. For most SAP environments (RAID, storage arrays, virtualizaton) 'noop' is the better choice.
With the new introduced multi-queue scheduler for block layer devices the recommended I/O scheduler is 'none' as an equivalent to 'noop' for single-queued block layer devices.

So IO_SCHEDULER can now contain a comma separated list of possible schedulers, which are checked from left to right. The first one which is available in */sys/block/<device>/queue/scheduler* will be used as new scheduler setting for the respective block device.
The selection per device is logged.
When set, **all** block devices on the system will be switched to one of the chosen schedulers.
Valid values can be found in */sys/block/<device>/queue/scheduler*.

NRREQ=*INT*

IO nr_requests specifies the maximum number of read and write requests that can be queued at one time. The default value is 128, which means that 128 read requests and 128 write requests can be queued before the next process to request a read or write is put to sleep.
When set, the number of requests for **all** block devices on the system will be switched to the chosen value

[cpu]

The section "[cpu]" manipulates files in */sys/devices/system/cpu/cpu**.
This section can only contain the following options:

energy_perf_bias=*STRING*

Energy Performance Bias EPB (applies to Intel-based systems only)
supported values are: **performance** (0), **normal** (6) and **powersave** (15)
The command 'cpupower set -b <value>' is used to set the value, if the system supports Intel's performance bias setting. See cpupower(1) and cpupower-set(1) for more information.
If system does not support Intel's performance bias setting - '**all:none**' is used in the column '*Actual*' of the verify table and the *footnote* '[1] setting is not supported by the system' is displayed.

When set as 'energy_perf_bias=<performance|normal|powersave> in the Note definition file, the value will be set for **all** available CPUs.
The command '**cpupower -c all set -b <value>**' or '**cpupower -c <cpu> set -b <value>**' is used to set the value.

governor=*STRING*

CPU Frequency/Voltage scaling (applies to Intel-based systems only)
The clock frequency and voltage of modern CPUs can scale, in order to save energy when there's less work to be done. However HANA as a high-performance database benefits from high CPU frequencies.
supported values are: **performance** (0), **normal** (6) and **powersave** (15)
The command 'cpupower frequency-set -g <value>' is used to set the value, if the value is a supported governor listed in */sys/devices/system/cpu/cpu*/cpufreq/scaling_governor*' See cpupower(1) and cpupower-frequency-set(1) for more information.
If the governor settings of all available CPUs are equal, '**all:<governor>**' is used in the column '*Actual*' of the verify table. If not, each CPU with its assigned governor is listed (e.g. cpu1:powersave cpu2:powersave cpu3:powersave cpu4:powersave cpu5:powersave cpu6:powersave

cpu7:powersave cpu0:performance)

When set as 'governor=<performance|powersave> in the Note definition file, the value will be set for **all** available CPUs.
The command '**cpupower -c all frequency-set -g <value>**' or '**cpupower -c <cpu> frequency-set -g <value>**' is used to set the value.

force_latency=*STRING*

force latency - configure C-States for lower latency (applies to Intel-based systems only)
Input is a string, which is internally treated as a decimal (not a hexadecimal) integer number representing a maximum response time in microseconds.
It is used to establish a latency upper limit by limiting the use of C-States (CPU idle or CPU latency states) to only those with an exit latency smaller than the value set here. That means only those states that require less than the requested number of microseconds to wake up are enabled, all the other C-States are disabled.
The files */sys/devices/system/cpu/cpu*/cpuidle/state*/latency* and */sys/devices/system/cpu/cpu*/cpuidle/state*/disable* are used to limit the C-States.

If system does not support force latency settings - '**all:none**' is used in the column '*Actual*' of the verify table and the *footnote* '[1] setting is not supported by the system' is displayed.

When set in the Note definition file for all available CPUs all CPU latency states with a value read from */sys/devices/system/cpu/cpu*/cpuidle/state*/latency* **>=** (higher than) the value from the Note definition file are disabled by writing '**1**' to */sys/devices/system/cpu/cpu*/cpuidle/state*/disable*

ATTENTION: not idling *at all* increases power consumption significantly and reduces the life span of the machine because of wear and tear. So do not use a too strict latency setting. For SAP HANA workloads a value of '**70**' microseconds (as a "light sleep") seems to be sufficient. And the impact on power consumption and life of the CPUs is less severe. But don't forget: The deeper the idle state, the larger is the exit latency.

[grub]

The section "[grub]" is checking kernel command line settings for grub. The values from the Note definition files are only checked against */proc/cmdline*. Changing the grub configuration is not supported by saptune.

Some of these values are set by saptune during runtime, so changing the grub configuration is possible but not needed.

This section can contain options like:

intel_idle.max_cstate=1 and **processor.max_cstate=1**

Configure C-States for lower latency in Linux (applies to Intel-based systems only) - see energy_perf_bias, governor and force latency in section [cpu]

numa_balancing=disable

Turn off autoNUMA balancing - see kernel.numa_balancing in section [sysctl]

transparent_hugepage=never

Disable transparent hugepages - see THP in section [vm]

[limits]

The section "[limits]" is dealing with ulimit settings for user login sessions in the pam_limits module. The settings will **NOT** be done in the central limits file */etc/security/limits.conf*. Instead there will be a **drop-in file** in */etc/security/limits.d* for each domain-item-type combination used in the Note definition file.

The drop-in file name syntax will be:
saptune-<domain>-<item>-<type>.conf

For more information and a description of the syntax and the needed fields please look at limits.conf(5).

This section has to contain the following option:

LIMITS=*STRING*

where STRING is a list of valid limit definitions separated by '**,**'
a valid limit definition contains the fields 'domain item type value' separated by one space
For more information about the syntax of valid limit definitions please refer to limits.conf(5) or the comment section of */etc/security/limits.conf*.
Note: The "@" sign in front of the domain name matches a group.

To leave **all** limits definitions of a Note definition file 'untouched' in the system, leave the **LIMITS** string in the **override file** of the Note definition file empty

To leave only **some** of the limits definitions of a Note definition file 'untouched' in the system, remove these limits definitions from the **LIMITS** string in the **override file** of the Note definition file.

[login]

The section "[login]" manipulates the behaviour of the systemd login manager.
This section can **only** contain the following option:

UserTasksMax=*STRING*

This option configures a parameter of the systemd login manager. It sets the maximum number of OS tasks each user may run concurrently. The behaviour of the systemd login manager was changed starting SLES12SP2 to prevent fork bomb attacks.

Recommended value is '**infinity**'.

If set, the drop-in file */etc/systemd/logind.conf.d/saptune-UserTasksMax.conf* is created and for all currently logged in users the maximum number of OS tasks each user may run concurrently is changed using the command '**systemctl --runtime set-property user-<uid>.slice TasksMax=<value>**'.
After creating the drop-in file the *systemd-logind.service* will be restarted.

ATTENTION: With this setting your system is vulnerable to fork bomb attacks

[mem]

The section "[mem]" manipulates the size of TMPFS (*/dev/shm*).

With the STD implementation, the SAP Extended Memory is no longer stored in the TMPFS (under /dev/shm). However, the TMPFS is required by the Virtual Machine Container (VMC). For this reason, we still recommend the same configuration of the TMPFS:
75% (RAM + Swap) is still recommended as the size.
This section can contain the following options:

ShmFileSystemSizeMB=*INT*

Use ShmFileSystemSizeMB to set an absolute value for your TMPFS.
If ShmFileSystemSizeMB is set to a value > 0, the setting for VSZ_TMPFS_PERCENT will be ignored and the size will NOT be calculated.
If ShmFileSystemSizeMB is set to '**0**' the size will be calculated using VSZ_TMPFS_PERCENT

VSZ_TMPFS_PERCENT=*INT*

Size of tmpfs mounted on */dev/shm* in percent of the virtual memory.
Depending on the size of the virtual memory (physical+swap) the value is calculated by (RAM + SWAP) * VSZ_TMPFS_PERCENT/100
If VSZ_TMPFS_PERCENT is set to '**0**', the value is calculated by (RAM + SWAP) * 75/100, as the default is 75.

[pagecache]

The section "[pagecache]" is dealing with the pagecache limit feature as described in SAP Note 1557506, which is only available on SLE12.

ATTENTION: The pagecache limit Note will **NOT** be part of any solution definition by default. As it is essential to configure this feature really carefully, you need to customize the Note definition file first to enable the feature and then you can apply the note settings manually. After that, the settings will be applied automatically during each startup of the system.
This section can contain the following options:

ENABLE_PAGECACHE_LIMIT=*yesno*

This defines whether pagecache limit feature should be enabled or not. It is a yesno value. By default it is set to **no**
Consider to enable pagecache limit feature if your SAP workloads cause frequent and excessive swapping activities. It is recommended to leave pagecache limit disabled if the system has low or no swap space.

vm.pagecache_limit_ignore_dirty=*INT*

Whether or not to ignore dirty memory when enforcing the pagecache limit.
If set to 0, dirty memory will be freed (written onto disk) when enforcing the pagecache limit.
If set to 1 (default), dirty memory will not be freed when enforcing the pagecache limit.
If set to 2 - a middle ground, some dirty memory will be freed when enforcing the limit.

OVERRIDE_PAGECACHE_LIMIT_MB=*INT*

When pagecache limit feature is enabled, the limit value is usually automatically calculated using the 'HANA formula', which means 2% of system memory is used as pagecache limit.
However, the value can be overridden if you set this parameter to the desired limit value.
To remove the override, set the parameter to empty string.

[reminder]

The section "[reminder]" contains important information and all settings of a SAP Note, which can not set by saptune.

This section is displayed at the end of the saptune options 'verify', 'simulate' and 'apply'. It will be highlighted with red color to get the attention of the customer.

[rpm]

The section "[rpm]" is checking rpm versions on the system. The values from the Note definition files are only checked against the installed rpm versions on the system. No other action is supported.
Package dependencies - if needed - are handled by the saptune package installation.

Syntax:
<rpm package name> <SLE Version> <rpm package version>
Add one line for each SLE version a package should be checked for, even if the package version is the same.
The SLE version has to be noted in the same format as the '**VERSION=**' entry in */etc/os-release*.

e.g

systemd 12-SP2 228-142.1
sapinit-systemd-compat 12 1.0-2.1
sapinit-systemd-compat 12-SP1 1.0-2.1
util-linux 12-SP1 2.25-22.1

Only the lines where the SLE version is matching the running system OS are checked and displayed during the 'verify' and 'simulate' option.
That means, if there is no matching SLE version for the running OS no rpm entries are listed during the 'verify' and 'simulate' operation.

[service]

The section "[service]" is dealing with starting and stopping services controlled by systemd.
The syntax for the entries are:

<servicename>=<start|stop>

Valid services are those listed by the command '*systemctl list-unit-files*'.
Valid values are '**start**' or '**stop**'.

Exceptions and Warnings:

For the service **uuidd.socket** only '**start**' is a valid value, because the uuidd.socket service is essential for a working SAP environment.

Concerning **sysstat.service** please be in mind: A running sysstat service can effect the system performance. But if there are real performance trouble with the SAP system, SAP service normally orders the sysstat reports collected in /var/log/sa.
See sar(1), sa2(8), sa1(8) for more information

If a service is enabled or disabled by default or admin choice, saptune will NOT disable or enable this service. It will only start/stop the service. If such a service is started by systemd during a system reboot **after** the start of tuned.service it will be possible that a service is stopped/running even if it was started/stopped by saptune.

[sysctl]

The section "[sysctl]" can be used to modify kernel parameters. The parameters available are those listed under /proc/sys/.
Please write the section keyword '[sysctl]' in the first line and add the desired tunables in 'sysctl.conf' syntax.

sysctl.parameter=*VALUE*

[vm]

The section "[vm]" manipulates */sys/kernel/mm* switches.
This section can to contain the following options:

THP=*STRING*

This option disables transparent hugepages by changing */sys/kernel/mm/transparent_hugepage/enabled*
Possible values are '**never**' to disable and '**always**' to enable.

KSM=*INT*

Kernel Samepage Merging (KSM). KSM allows for an application to register with the kernel so as to have its memory pages merged with other processes that also register to have their pages merged. For KVM the KSM mechanism allows for guest virtual machines to share pages with each other. In today's environment where many of the guest operating systems like XEN, KVM are similar and are running on same host machine, this can result in significant memory savings, the default value is set to 0.

FILES

/usr/share/saptune/notes

here you can find examples how to set the 'parameter value' pairs of the available sections.
But please do not change the files located here. You will lose all your changes during a saptune package update. Use an override or extra file for your changes as described in saptune_v2(8).

SEE ALSO

saptune-migrate(7) saptune(8) saptune_v1(8) saptune_v2(8) tuned(8) tuned-adm(8)

AUTHOR

Soeren Schmidt <soeren.schmidt@suse.com>, Angela Briel <abriel@suse.com>