



Red Hat Enterprise Linux 6

Guide de gestion de l'alimentation

(Bêta) Gestion de la consommation d'énergie sur Red Hat Enterprise Linux 6
Édition 0.2

Last Updated: 2017-10-12

DRAFT

DRAFT

Red Hat Enterprise Linux 6 Guide de gestion de l'alimentation

(Bêta) Gestion de la consommation d'énergie sur Red Hat Enterprise Linux 6
Édition 0.2

Don Domingo
Red Hat Engineering Content Services

Rüdiger Landmann
Red Hat Engineering Content Services
r.landmann@redhat.com

Red Hat Inc.



Notice légale

Copyright © 2010 Red Hat Inc..

This document is licensed by Red Hat under the [Creative Commons Attribution-ShareAlike 3.0 Unported License](#). If you distribute this document, or a modified version of it, you must provide attribution to Red Hat, Inc. and provide a link to the original. If the document is modified, all Red Hat trademarks must be removed.

Red Hat, as the licensor of this document, waives the right to enforce, and agrees not to assert, Section 4d of CC-BY-SA to the fullest extent permitted by applicable law.

Red Hat, Red Hat Enterprise Linux, the Shadowman logo, JBoss, OpenShift, Fedora, the Infinity logo, and RHCE are trademarks of Red Hat, Inc., registered in the United States and other countries.

Linux ® is the registered trademark of Linus Torvalds in the United States and other countries.

Java ® is a registered trademark of Oracle and/or its affiliates.

XFS ® is a trademark of Silicon Graphics International Corp. or its subsidiaries in the United States and/or other countries.

MySQL ® is a registered trademark of MySQL AB in the United States, the European Union and other countries.

Node.js ® is an official trademark of Joyent. Red Hat Software Collections is not formally related to or endorsed by the official Joyent Node.js open source or commercial project.

The OpenStack ® Word Mark and OpenStack logo are either registered trademarks/service marks or trademarks/service marks of the OpenStack Foundation, in the United States and other countries and are used with the OpenStack Foundation's permission. We are not affiliated with, endorsed or sponsored by the OpenStack Foundation, or the OpenStack community.

All other trademarks are the property of their respective owners.

Résumé

Ce document explique comment gérer la consommation d'énergie sur les systèmes Red Hat Enterprise Linux 6 de manière effective. Les sections suivantes discutent des différentes techniques permettant d'abaisser la consommation d'énergie (pour les ordinateurs portables et les serveurs), et de la manière dont chaque technique affectera la performance globale de votre système. Ce document est en cours de développement, est sujet à de nombreux changements, et vous est fourni en tant qu'aperçu. Le contenu et les instructions qui s'y trouvent ne doivent pas être considérés comme complets, et devraient être utilisés avec précautions.

Table des matières

CHAPITRE 1. VUE D'ENSEMBLE	3
1.1. IMPORTANCE DE LA GESTION DE L'ALIMENTATION	3
1.2. BASES DE LA GESTION DE L'ALIMENTATION	4
CHAPITRE 2. AUDIT ET ANALYSE DE LA GESTION DE L'ALIMENTATION	7
2.1. APERÇU DES AUDITS ET DES ANALYSES	7
2.2. POWERTOP	7
2.3. DISKDEVSTAT ET NETDEVSTAT	9
2.4. BATTERY LIFE TOOL KIT	13
2.5. TUNED ET KTUNE	14
2.5.1. Le fichier tuned.conf	16
2.5.2. Tuned-adm	17
2.6. DEVICEKIT-POWER ET DEVKIT-POWER	19
2.7. GESTIONNAIRE D'ALIMENTATION GNOME	20
2.8. AUTRES MANIÈRES D'AUDITER	21
CHAPITRE 3. INFRASTRUCTURE ET MÉCANIQUE PRINCIPALE	22
3.1. ÉTATS INACTIFS DU CPU	22
3.2. UTILISATION DES GOUVERNEURS CPUFREQ	22
3.2.1. Types de gouverneurs CPUfreq	23
3.2.2. Installation de CPUfreq	24
3.2.3. Paramétrer la politique CPUfreq et la vitesse	25
3.3. SUSPENSION ET REPRISE	26
3.4. NOYAU TICKLESS	27
3.5. GESTION DE L'ALIMENTATION EN ÉTAT ACTIF	27
3.6. AGGRESSIVE LINK POWER MANAGEMENT	28
3.7. OPTIMISATION D'ACCÈS AU DISQUE RELATIME	29
3.8. LIMITATION DE LA CONSOMMATION	29
3.9. GESTION AMÉLIORÉE DE L'ALIMENTATION DES GRAPHISMES	30
3.10. RFKILL	31
3.11. OPTIMISATIONS DANS L'ESPACE UTILISATEUR	32
CHAPITRE 4. CAS D'UTILISATION	34
4.1. EXEMPLE – SERVEUR	34
4.2. EXEMPLE – ORDINATEUR PORTABLE	35
ANNEXE A. ASTUCES POUR LES DÉVELOPPEURS	38
A.1. UTILISATION DE THREADS	38
A.2. RÉVEILS	39
A.3. FSYNC	40
ANNEXE B. HISTORIQUE DES RÉVISIONS	41

DRAFT

DRAFT

CHAPITRE 1. VUE D'ENSEMBLE

La gestion de l'alimentation fut l'un des points principaux sur lequel nous nous sommes concentrés lors des améliorations apportées à Red Hat Enterprise Linux 6. Limiter la consommation d'électricité des systèmes informatiques est l'un des aspects les plus importants de l'*IT verte* (informatique respectueuse de l'environnement), un ensemble de considérations qui englobe aussi l'utilisation de matériaux recyclables, l'impact sur l'environnement de la production de matériel, et la prise de conscience de l'environnement dans la conception et le déploiement des systèmes. Dans ce document, nous vous offrons des conseils et informations concernant la gestion de l'alimentation de vos systèmes exécutant Red Hat Enterprise Linux 6.

1.1. IMPORTANCE DE LA GESTION DE L'ALIMENTATION

Au coeur de la gestion de l'alimentation figure la compréhension de l'optimisation efficace de la consommation d'énergie pour chaque composant d'un système. Ceci englobe l'étude des différentes tâches effectuées par votre système et la configuration de chaque composant pour vous assurer que sa performance est adéquate pour la tâche à entreprendre.

Les principales motivations pour la gestion de l'alimentation sont :

- la réduction de la consommation d'énergie en général dans le but de réduire les coûts

Une gestion de l'alimentation correcte résulte en :

- une réduction de chaleur pour les serveurs et les centres de calcul
- des coûts secondaires réduits, y compris le refroidissement, l'espace, les câbles, les générateurs, et les *systèmes d'alimentation sans coupure*(UPS)
- une augmentation de l'autonomie de la batterie pour les ordinateurs portables
- une baisse de la production de dioxyde de carbone
- conformité avec la réglementation du gouvernement ou avec les exigences légales concernant l'*IT verte*, par exemple Energy Star
- conformité avec la ligne de conduite de la compagnie sur les nouveaux systèmes

Par définition, la baisse de consommation d'énergie pour un composant spécifique (ou du système en tant que tout) conduira à une baisse de chaleur, et évidemment, à une baisse de performance. Ainsi, vous devriez étudier et tester minutieusement les baisses de performance résultant des configurations que vous avez effectué, tout particulièrement pour les systèmes menant à bien des missions critiques.

En étudiant les différentes tâches menées par votre système, et en configurant chaque composant afin de vous assurer que leur performance est suffisante pour la tâche, vous pouvez économiser de l'énergie, générer moins de chaleur, et optimiser l'autonomie de la batterie pour les ordinateurs portables. La plupart des principes pour l'analyse et le paramétrage d'un système par rapport à sa consommation d'électricité sont similaires à ceux du paramétrage de la performance. D'une certaine manière, la gestion de l'alimentation et le paramétrage de la performance sont des approches à la configuration d'un système opposées, car habituellement les systèmes sont optimisés soit vers la performance, soit vers une meilleure gestion de l'alimentation. Ce manuel décrit les outils fournis par Red Hat, ainsi que les techniques que nous avons développé afin de vous assister dans ce processus.

Red Hat Enterprise Linux 6 contient déjà de nombreuses fonctionnalités de gestion de l'alimentation qui sont activées par défaut. Celles-ci ont été choisies minutieusement afin de ne pas avoir d'impact sur la performance d'une utilisation typique d'un serveur ou d'un bureau. Cependant, dans certains cas

très spécifiques où le débit maximal, une latence minimale, ou quand la performance CPU la plus importante est un impératif, une révision de ces valeurs par défaut est requise.

Posez-vous quelques questions afin de décider si vous devriez optimiser vos machines à l'aide des techniques décrites dans ce document :

Q : Est-ce que je dois vraiment optimiser ?

R : L'importance de l'optimisation de l'alimentation dépend de si votre compagnie donne des indications à suivre ou de s'il existe une réglementation à observer.

Q : À quel point devrais-je optimiser ?

R : Certaines des techniques que nous vous présentons ne requièrent pas que vous procédiez à la totalité du processus d'audit et d'analyse détaillée de votre machine, au contraire elles vous fourniront un ensemble d'optimisations générales qui améliorent l'utilisation de l'alimentation. Celles-ci, bien entendu, ne seront pas aussi bonnes qu'avec un système audité et optimisé manuellement, mais elles fourniront quand même un bon compromis.

Q : Est-ce que l'optimisation va réduire la performance du système à un niveau inacceptable ?

R : La plupart des techniques décrites dans ce document auront un impact qui se ressentira sur la performance de votre système. Si vous choisissez d'implémenter une gestion de l'alimentation autre que celle installée par défaut dans Red Hat Enterprise Linux 6, vous devriez alors surveiller la performance du système après l'optimisation, puis décidez si la perte de performance est acceptable ou non.

Q : Le temps pris et les ressources utilisées pour optimiser le système sont-ils plus importants que les gains à obtenir ?

R : L'optimisation manuelle d'un système unique en suivant la totalité du processus n'en vaut pas la peine car le temps et le coût pris sont bien plus importants que les bénéfices à obtenir, et ce sur la durée entière du cycle de vie de la machine en question. Cependant, sur 10000 systèmes informatiques dans vos bureaux utilisant la même configuration et le même paramétrage, la création d'un paramétrage optimisé et s'appliquant à ces 10000 machines serait très certainement une bonne idée.

les sections suivantes vous expliqueront en quoi la performance d'un matériel optimal peut bénéficier à votre système en termes de consommation d'énergie.

1.2. BASES DE LA GESTION DE L'ALIMENTATION

La gestion d'une alimentation effective est construite sur les principes suivants :

Un CPU inactif ne devrait se réveiller que lorsqu'il est nécessité

Le noyau de Red Hat Enterprise Linux 5 utilisait un minuteur périodique pour chaque CPU. Ce minuteur empêchait le CPU de réellement devenir inactif en l'obligeant à calculer chaque événement de minuteur (ce qui arrive toutes les quelques millisecondes, en fonction du paramétrage), sans tenir compte de quels processus étaient (ou n'étaient pas) en cours d'exécution. Une grande partie de la gestion de l'alimentation implique la réduction de la fréquence des réveils du CPU.

À cause de ceci, le noyau Linux dans Red Hat Enterprise Linux 6 élimine le minuteur périodique : par conséquent, l'état inactif d'un CPU est maintenant *tickless*. Ceci empêche le CPU de consommer de l'énergie non-nécessaire lorsqu'il est inactif. Cependant, les bénéfices de cette fonctionnalité peuvent être perdus si votre système possède des applications créant des événements de minuteur non-nécessaires. Les événements d'interrogation (tels que les vérifications de changements de volume, le mouvement de la souris, etc.) sont des exemples de tels événements.

Red Hat Enterprise Linux 6 inclut des outils avec lesquels vous pourrez identifier et auditer des applications sur la base de l'utilisation de leur CPU. Reportez-vous à [Chapitre 2, Audit et analyse de la gestion de l'alimentation](#) pour plus de détails.

Le matériel et les périphériques inutilisés devraient être complètement désactivés

Ceci est particulièrement vrai pour les périphériques qui possèdent des pièces mobiles (par exemple, des disques durs). De plus, certaines applications peuvent laisser un périphérique qui est activé mais non-utilisé sur un état « ouvert » (de l'anglais, « open ») ; lorsque ceci se produit, le noyau assume que le périphérique est en cours d'utilisation, ce qui peut empêcher le périphérique de se mettre dans un état d'économie d'énergie.

Une faible activité devrait signifier une faible puissance

Dans de nombreux cas, ceci dépend d'un matériel moderne et d'une configuration BIOS correcte. Les composants de système plus anciens ne prennent pas en charge certaines des nouvelles fonctionnalités prises en charge dans Red Hat Enterprise Linux 6. Assurez-vous de bien utiliser le microprogramme le plus récent dans vos systèmes et que les fonctionnalités de gestion de l'alimentation dans les sections de gestion de l'alimentation ou de configuration de périphériques sont activées. Ci-dessous figurent quelques unes des fonctionnalités à rechercher :

- SpeedStep
- PowerNow!
- Cool'n'Quiet
- ACPI (C state)
- Smart

Si votre matériel prend ces fonctionnalités en charge et qu'elles sont activées dans le BIOS, Red Hat Enterprise Linux 6 les utilisera par défaut.

Les différentes formes des états de CPU et leurs effets

Les CPUs modernes ainsi que l'*ACPI* (de l'anglais, Advanced Configuration and Power Interface) offrent différents états d'alimentation. Les trois différents états sont :

- Veille (C-states)
- Fréquence (P-states)
- Sortie de chaleur (T-states ou "états thermaux")

Un CPU exécuté sur l'état de veille le plus bas possible consomme un minimum de watts, mais prend aussi considérablement plus de temps à sortir de cet état lorsque nécessaire. Dans de rares cas, ceci peut faire que le CPU ne doive se réveiller immédiatement après être allé en veille. Cette situation résulte en un CPU constamment occupé et donc en une plus grande perte des économies potentielles d'énergie que si un autre état avait été utilisé.

Une machine éteinte consomme le moins d'énergie

Aussi évident que cela puisse paraître, la meilleure manière d'économiser de l'énergie est d'éteindre les systèmes. Par exemple, votre compagnie peut développer une culture d'entreprise autour de « l'IT verte » avec une ligne de conduite dans laquelle il vous faudrait éteindre les machines pendant les coupures déjeuner, ou lorsque vous rentrez chez vous. Vous pourriez aussi consolider plusieurs serveurs physique en un grand serveur et les virtualiser à l'aide de la technologie de virtualisation que nous vous fournissons avec Red Hat Enterprise Linux 6.



CHAPITRE 2. AUDIT ET ANALYSE DE LA GESTION DE L'ALIMENTATION

2.1. APERÇU DES AUDITS ET DES ANALYSES

L'audit, analyse, et paramétrage détaillé et manuel d'un système unique est habituellement l'exception, car le temps et le coût que cela prend dépassent les bénéfices obtenus par ces derniers. Cependant, effectuer ces tâches une seule fois pour un grand nombre de systèmes pratiquement identiques, pour lesquels vous pourrez réutiliser les mêmes paramètres pour tous les systèmes peut être très utile. Par exemple, prenez le déploiement de milliers de systèmes de bureau ou un cluster HPC où les machines seraient quasiment identiques. Une autre raison de réaliser un audit et une analyse serait de fournir une base pour une comparaison dans laquelle vous identifieriez les régressions ou changements dans le comportement du système dans le futur. Les résultats de cette analyse peuvent se révéler très utiles lorsque des mises à jour du matériel, du BIOS, ou de logiciels se produisent de manière régulière et que vous voulez éviter toute surprise quant à la consommation d'électricité. En général, un audit et une analyse minutieuse vous offriront une bien meilleure idée de ce qu'il se passe sur un système en particulier.

Auditer et analyser un système sur sa consommation d'énergie est relativement difficile, même avec les systèmes les plus modernes. La plupart des systèmes n'offrent pas les moyens nécessaires de mesurer l'énergie utilisée via software. Il existe pourtant des exceptions : la console de gestion ILO des systèmes de serveurs de Hewlett Packard offre un module de gestion de l'alimentation accessible depuis le web. IBM offre une solution similaire dans leur module de gestion de l'alimentation BladeCenter. Sur certains systèmes Dell, l'assistant « IT Assistant » offre aussi des possibilités de surveillance de l'alimentation. Il est possible que d'autres fournisseurs puissent fournir des solutions similaires pour leurs plateformes serveur, mais comme on peut l'observer, il n'existe pas de solution uniforme prise en charge par tous les fournisseurs. Si votre système ne possède pas de mécanisme intégré pour mesurer la consommation d'électricité, il existe quelques autres solutions. Vous pourriez installer un bloc d'alimentation spécifique sur votre système, celui-ci offrirait des informations sur la consommation d'électricité via USB. Le bloc d'alimentation PC Odin GT Gygabyte 550 W en est un exemple, et le logiciel pour en lire les valeurs sous Linux est disponible extérieurement sur <http://mgmt.sth.sze.hu/~andras/dev/gopsu/>. En dernier recours, certains wattmètres externes, comme le Watts up? PRO possède une prise USB.

Les mesures directes de consommation d'électricité sont souvent nécessaires afin de maximiser les économies autant que possible. Heureusement, il existe d'autres moyens pour mesurer si les changements sont en cours ou comment se comporte le système. Ce chapitre décrit les outils y étant nécessaires.

2.2. POWERTOP

Le lancement du nouveau noyau sans tic (de l'anglais, « tickless kernel ») dans Red Hat Enterprise Linux 6 (reportez-vous à la [Section 3.4, « Noyau Tickless »](#)) permet au CPU d'entrer en état inactif plus fréquemment, en réduisant la consommation d'électricité et en améliorant la gestion de l'alimentation. Le nouvel outil **PowerTOP** identifie les composants spécifiques des applications du noyau et de l'espace utilisateur qui réveillent fréquemment le CPU. **PowerTOP** était utilisé dans le développement afin de procéder à des audits décrits dans la [Section 3.11, « Optimisations dans l'espace utilisateur »](#) qui ont conduit des applications à être ajustées dans cette version, réduisant ainsi le nombre de réveils CPU non-nécessaires par un facteur 10.

Installez **PowerTOP** avec cette commande :

```
yum install powertop
```

Exécutez **PowerTOP** avec cette commande :

```
power top
```

Remarquez que vous devrez exécuter **PowerTOP** avec des privilèges root afin de permettre à l'application de se rendre utile.

Lorsqu'il est exécuté, **PowerTOP** rassemble des statistiques du système et vous fournit la liste des composants qui envoient des réveils CPU le plus fréquemment. **PowerTOP** vous fait aussi des suggestions quant au paramétrage du système pour obtenir une consommation plus basse. Ces suggestions apparaissent au bas de l'écran, et vous spécifient une touche à saisir pour accepter la suggestion de **PowerTOP**. Comme **PowerTOP** se réactualise périodiquement, des suggestions supplémentaires apparaissent. Dans la [Figure 2.1, « Opération de PowerTOP »](#), remarquez la suggestion « `increase the VM dirty writeback time` », et la touche (W) sur laquelle appuyer pour accepter la suggestion.

Lorsqu'il est exécuté, **PowerTOP** collecte des statistiques du système et vous présente plusieurs listes d'informations importantes. Tout en haut figure la liste vous indiquant depuis combien de temps les coeurs de CPU ont été dans chaque états C et P disponible. Plus le CPU reste longtemps dans des états C ou P hauts (C4 étant plus haut que C3), mieux c'est ; de plus, cela indique aussi que le système est paramétré dans le but d'utiliser le CPU de manière plus efficace. L'idéal étant que le CPU soit à 90% ou plus dans l'état C ou P le plus haut lorsque le système est inactif.

Le deuxième élément est un sommaire des réveils par seconde sur la machine. Le nombre de réveils par seconde présent une mesure correcte de la manière dont les services ou périphériques et les pilotes du noyau s'exécutent en relation à la consommation de l'alimentation sur votre système. Plus il y a de réveils par secondes, plus l'énergie est consommée ; dans le cas présent, moins il y aura de réveils, meilleur ce sera.

Si disponible, **PowerTOP** offre une estimation de la consommation actuelle du système. Attendez-vous à ce que **PowerTOP** rapporte ce chiffre sur les ordinateurs portables lorsqu'ils sont en mode batterie.

Toute estimation disponible de la consommation de l'alimentation est suivie par une liste détaillée des composants qui envoient des réveils au CPU le plus fréquemment. En haut de la liste figurent les composants que vous devriez étudier de plus près afin d'optimiser votre système de manière à réduire la consommation. Si ceux-ci se révèlent être des composants du noyau, (ce qui est indiqué lorsque le composant se trouve dans <>), alors les réveils sont souvent associés au pilote spécifique qui les provoque. Paramétrer les pilotes requiert habituellement des modifications du noyau allant au-delà de la portée de ce document. Cependant, les processus de l'espace utilisateur qui envoient des réveils sont plus faciles à gérer. Commencez par identifier si ce service ou cette application doit réellement s'exécuter sur le système ou non. Si non, désactivez-le (ou désactivez-la). Pour désactiver un service de manière permanente, exécutez :

```
chkconfig nomduservice off
```

Si vous avez besoin de plus de détails sur la fonction exacte du le compo, exécutez :ant

```
ps -awux | grep nomducomposant  
strace -p idprocessus
```

Si le suivi à l'air de se répéter, vous avez probablement trouvé une attente active (de l'anglais, busy loop). Corriger ceci nécessiterait une modification du code du composant, ce qui, une fois de plus, se trouve au-delà de la portée de ce document.

Finalement, **PowerTOP** fait aussi des suggestions pour paramétrer le système de manière à réduire la

consommation de l'alimentation. Ces suggestions apparaissent au bas de l'écran, ainsi qu'une touche sur laquelle appuyer afin d'accepter la suggestion de **PowerTOP**. Comme **PowerTOP** se réactualise de manière périodique, des suggestions supplémentaires apparaissent. Dans la [Figure 2.1, « Opération de PowerTOP »](#), remarquez la suggestion `increase the VM dirty writeback time`, ainsi que la touche **(W)** sur laquelle appuyer pour accepter la suggestion. Ces changements ne seront actifs que lors du prochain redémarrage. Pour vous aider à rendre ces changements permanents, **PowerTOP** affiche la commande exacte qu'il utilise pour effectuer cette optimisation. Ajoutez la commande à votre fichier `/etc/rc.local` à l'aide de votre éditeur de texte préféré afin que ce changement soit effectué à chaque fois que l'ordinateur est démarré.

```
PowerTOP version 1.11 (C) 2007 Intel Corporation

Cn          Avg residency      P-states (frequencies)
C0 (cpu running)  ( 4.4%)          2.81 Ghz    3.2%
polling      0.1ms ( 0.0%)    2.81 Ghz    0.2%
C1 mwait     0.0ms ( 0.0%)    2.14 Ghz    0.1%
C2 mwait     0.5ms ( 1.1%)    1.60 Ghz    0.4%
C4 mwait     4.3ms (94.5%)    800 Mhz     96.2%

Wakeups-from-idle per second : 245.5 interval: 15.0s
no ACPI power usage estimate available

Top causes for wakeups:
 38.3% (163.7) <kernel core> : hrtimer_start_range_ns (tick_sched_timer)
  8.8% ( 37.8) <interrupt> : iwlgagn
  8.6% ( 36.9) <kernel IPI> : Rescheduling interrupts
  7.9% ( 33.9) <interrupt> : extra timer interrupt
  7.9% ( 33.7) firefox : hrtimer_start_range_ns (hrtimer_wakeup)
  4.6% ( 19.9) popfile.pl : hrtimer_start_range_ns (hrtimer_wakeup)
  3.2% ( 13.8) <kernel core> : hrtimer_start (tick_sched_timer)
  2.7% ( 11.7) <interrupt> : i915
  2.6% ( 11.2) <interrupt> : ahci
  2.2% (  9.5) <interrupt> : ehci_hcd:usb1
  2.2% (  9.5) USB device 1-5.1.2 : Microsoft 3-Button Mouse with IntelliEye(TM) (Microsoft)
  2.1% (  9.0) <kernel core> : __mod_timer (ehci_watchdog)
  1.5% (  6.5) thunderbird-bin : hrtimer_start_range_ns (hrtimer_wakeup)
  1.3% (  5.5) simpres.b.in : hrtimer_start_range_ns (hrtimer_wakeup)
  1.3% (  5.5) plasma-desktop : hrtimer_start_range_ns (hrtimer_wakeup)
  1.2% (  5.3) <interrupt> : eth0
  0.9% (  4.0) <kernel core> : __mod_timer (rh_timer_func)
  0.2% (  1.0) klipper : hrtimer_start_range_ns (hrtimer_wakeup)
  0.2% (  1.0) httpd : hrtimer_start_range_ns (hrtimer_wakeup)
  0.2% (  0.9) konversation : hrtimer_start_range_ns (hrtimer_wakeup)

Suggestion: increase the VM dirty writeback time from 5.00 to 15 seconds with:
echo 1500 > /proc/sys/vm/dirty_writeback_centisecs
This wakes the disk up less frequently for background VM activity

Q - Quit R - Refresh W - Increase Writeback time
```

Figure 2.1. Opération de PowerTOP

Le site web *Less Watts* publie une liste d'applications que **PowerTOP** a identifié comme gardant les CPUs actifs. Reportez-vous à <http://www.lesswatts.org/projects/powertop/known.php>.

2.3. DISKDEVSTAT ET NETDEVSTAT

`Diskdevstat` et `netdevstat` sont des outils **SystemTap** qui collectent des informations détaillées sur les activités disque et réseau de toutes les applications exécutées sur un système. Ces outils ont été inspirés par **PowerTOP**, qui montre le nombre de réveils CPU par seconde de chaque application (reportez-vous à [Section 2.2, « PowerTOP »](#)). Les statistiques collectées par ces outils vous permettent d'identifier les applications qui utilisent de l'énergie inutilement avec beaucoup de petites opérations d'E/S, au lieu de peu d'opérations de grande taille. Les autres outils de surveillance ne mesurant que la vitesse de transfert n'aident pas à identifier ce type d'utilisation.

Installez ces outils avec **SystemTap** à l'aide de la commande :

```
yum install systemtap tuned-utils kernel-debuginfo
```

Exécutez ces outils avec la commande :

```
diskdevstat
```

ou avec la commande :

```
netdevstat
```

Ces deux commandes peuvent inclure jusqu'à trois paramètres, comme suit :

```
diskdevstat update_interval total_duration display_histogram
```

```
netdevstat update_interval total_duration display_histogram
```

update_interval

Le temps en secondes entre les mises à jour de l'affichage. Par défaut : 5

total_duration

Le temps en secondes pour la totalité de l'exécution. Par défaut : 86400 (1 jour)

display_histogram

Indiquez s'il faut inclure toutes les données collectées dans l'histogramme à la fin de l'exécution.

La sortie ressemble à celle de **PowerTOP**. Ci-dessous figure un exemple de sortie d'une plus longue exécution de **diskdevstat** sur un système Fedora 10 exécutant KDE 4.2 :

PID	UID	DEV	WRITE_CNT	WRITE_MIN	WRITE_MAX	WRITE_AVG	READ_CNT
READ_MIN	READ_MAX	READ_AVG	COMMAND				
2789	2903	sda1	854	0.000	120.000	39.836	0
0.000	0.000	0.000	plasma				
15494	0	sda1	0	0.000	0.000	0.000	758
0.000	0.012	0.000	0logwatch				
15520	0	sda1	0	0.000	0.000	0.000	140
0.000	0.009	0.000	perl				
15549	0	sda1	0	0.000	0.000	0.000	140
0.000	0.009	0.000	perl				
15585	0	sda1	0	0.000	0.000	0.000	108
0.001	0.002	0.000	perl				
2573	0	sda1	63	0.033	3600.015	515.226	0
0.000	0.000	0.000	auditd				
15429	0	sda1	0	0.000	0.000	0.000	62
0.009	0.009	0.000	crond				
15379	0	sda1	0	0.000	0.000	0.000	62
0.008	0.008	0.000	crond				
15473	0	sda1	0	0.000	0.000	0.000	62
0.008	0.008	0.000	crond				
15415	0	sda1	0	0.000	0.000	0.000	62
0.008	0.008	0.000	crond				
15433	0	sda1	0	0.000	0.000	0.000	62
0.008	0.008	0.000	crond				

15425	0 sda1	0	0.000	0.000	0.000	62
0.007	0.007	0.000 crond				
15375	0 sda1	0	0.000	0.000	0.000	62
0.008	0.008	0.000 crond				
15477	0 sda1	0	0.000	0.000	0.000	62
0.007	0.007	0.000 crond				
15469	0 sda1	0	0.000	0.000	0.000	62
0.007	0.007	0.000 crond				
15419	0 sda1	0	0.000	0.000	0.000	62
0.008	0.008	0.000 crond				
15481	0 sda1	0	0.000	0.000	0.000	61
0.000	0.001	0.000 crond				
15355	0 sda1	0	0.000	0.000	0.000	37
0.000	0.014	0.001 laptop_mode				
2153	0 sda1	26	0.003	3600.029	1290.730	0
0.000	0.000	0.000 rsyslogd				
15575	0 sda1	0	0.000	0.000	0.000	16
0.000	0.000	0.000 cat				
15581	0 sda1	0	0.000	0.000	0.000	12
0.001	0.002	0.000 perl				
15582	0 sda1	0	0.000	0.000	0.000	12
0.001	0.002	0.000 perl				
15579	0 sda1	0	0.000	0.000	0.000	12
0.000	0.001	0.000 perl				
15580	0 sda1	0	0.000	0.000	0.000	12
0.001	0.001	0.000 perl				
15354	0 sda1	0	0.000	0.000	0.000	12
0.000	0.170	0.014 sh				
15584	0 sda1	0	0.000	0.000	0.000	12
0.001	0.002	0.000 perl				
15548	0 sda1	0	0.000	0.000	0.000	12
0.001	0.014	0.001 perl				
15577	0 sda1	0	0.000	0.000	0.000	12
0.001	0.003	0.000 perl				
15519	0 sda1	0	0.000	0.000	0.000	12
0.001	0.005	0.000 perl				
15578	0 sda1	0	0.000	0.000	0.000	12
0.001	0.001	0.000 perl				
15583	0 sda1	0	0.000	0.000	0.000	12
0.001	0.001	0.000 perl				
15547	0 sda1	0	0.000	0.000	0.000	11
0.000	0.002	0.000 perl				
15576	0 sda1	0	0.000	0.000	0.000	11
0.001	0.001	0.000 perl				
15518	0 sda1	0	0.000	0.000	0.000	11
0.000	0.001	0.000 perl				
15354	0 sda1	0	0.000	0.000	0.000	10
0.053	0.053	0.005 lm_lid.sh				

Les colonnes sont :

PID

l'ID du processus de l'application

UID

l'ID utilisateur sous lequel l'application est exécutée

DEV

le périphérique sur lequel les E/S ont eu lieu

WRITE_CNT

le nombre total des opérations d'écriture

WRITE_MIN

le temps le moins long pris pour deux écritures consécutives (en secondes)

WRITE_MAX

le temps le plus long pris pour deux écritures consécutives (en secondes)

WRITE_AVG

le temps moyen pris pour deux écritures consécutives (en secondes)

READ_CNT

le nombre total des opérations de lecture

READ_MIN

le temps le moins long pris pour deux lectures consécutives (en secondes)

READ_MAX

le temps le plus long pris pour deux lectures consécutives (en secondes)

READ_AVG

le temps moyen pris pour deux lectures consécutives (en secondes)

COMMAND

le nom du processus

Dans cet exemple, trois applications se démarquent de manière évidente :

PID	UID	DEV	WRITE_CNT	WRITE_MIN	WRITE_MAX	WRITE_AVG	READ_CNT
READ_MIN	READ_MAX	READ_AVG	COMMAND				
2789	2903	sda1	854	0.000	120.000	39.836	0
0.000	0.000	0.000	plasma				
2573	0	sda1	63	0.033	3600.015	515.226	0
0.000	0.000	0.000	auditd				
2153	0	sda1	26	0.003	3600.029	1290.730	0
0.000	0.000	0.000	rsyslogd				

Ces trois applications ont un **WRITE_CNT** plus important que 0, ce qui veut dire qu'elles ont effectuées quelque sorte d'écriture pendant la mesure. De celles-ci, **plasma** fut la pire et de loin : cette application a effectuée la plupart des opérations d'écriture, et bien sûr le temps moyen entre les écritures était le plus bas. **Plasma** serait donc l'application idéale à étudier si vous vous sentiez concerné par les applications inefficaces au niveau de la consommation d'énergie.

Utilisez les commandes **strace** et **ltrace** pour examiner les applications de plus près en retraçant tous les appels système de l'ID du processus. Dans l'exemple présent, vous pourriez exécuter :

```
strace -p 2789
```

Dans cet exemple, la sortie de **strace** contient un schéma se répétant toutes les 45 secondes qui ouvre le fichier cache de l'icône KDE de l'utilisateur pour une écriture, suivi par la fermeture immédiate du fichier. Ceci a nécessairement conduit à une écriture physique sur le disque dur puisque les métadonnées du fichier (plus spécifiquement l'heure de la modification) avaient changées. Le correctif final servait à prévenir ces appels inutiles lorsqu'il n'y avait pas eu de mises à jour des icônes.

2.4. BATTERY LIFE TOOL KIT

Red Hat Enterprise Linux 6 présente le **Battery Life Tool Kit** (en français, kit de ressources de l'autonomie de la batterie, ou BLTK), une suite de tests qui simule et analyse l'autonomie et la performance de la batterie. BLTK achève ceci en procédant à une série de tâches simulant des groupes d'utilisateurs spécifiques, tels que des développeurs, ou employés de bureau et en rapportant les résultats. Principalement développé pour tester la performance des ordinateurs portables, BLTK peut aussi être utilisé pour créer des rapports sur la performance des ordinateurs de bureau lorsque lancé avec **-a**.

BLTK vous permet de générer des charges de travail reproductibles comparables à une utilisation réelle d'une machine. Par exemple, la charge de travail **office** écrit un texte, le corrige, et refera la même chose sur une feuille de calcul. L'exécution de BLTK, combinée avec **PowerTOP**, ou tout autre outil d'audit ou d'analyse, vous permettra de vérifier si l'optimisation que vous avez effectuée aurait un effet lorsque la machine est utilisée au lieu de lorsqu'elle n'est que simplement inactive. Comme vous pouvez exécuter la même charge de travail de multiples fois avec différents paramètres, vous pourrez comparer les résultats de ceux-ci.

Installez BLTK avec la commande :

```
yum install bltk
```

Exécutez BLTK avec la commande :

```
bltk workload options
```

Par exemple, pour exécuter la charge de travail **inactive** pendant 120 secondes :

```
bltk -I -T 120
```

Les charges de travail disponibles par défaut sont :

-I, --idle

système est inactif, pour utiliser comme ligne de base pour comparer avec d'autres charges de travail

-R, --reader

simule la lecture de documents (par défaut avec **Firefox**)

-P, --player

simule la lecture de fichiers multimédia depuis un lecteur CD ou DVD (par défaut avec **mplayer**)

-O, --office

simule la modification de documents avec la suite bureautique **OpenOffice.org**

D'autres options vous permettent de spécifier :

-a, --ac-ignore

ignore si l'alimentation secteur est disponible (nécessaire pour l'utilisation d'ordinateurs de bureau)

-T nombre_de_secondes, --time nombre_de_secondes

le temps (en secondes) pris pour l'exécution du test ; utilisez cette option avec la charge de travail **inactive**

-F nom_de_fichier, --file nom_de_fichier

spécifie un fichier devant être utilisé par une charge de travail en particulier, par exemple un fichier pour la charge de travail **player** pour lire au lieu d'accéder au lecteur CD ou DVD.

-W application, --prog application

spécifie une application devant être utilisée par une charge de travail, par exemple un navigateur autre que **Firefox** pour la charge de travail **reader**

BLTK prend en charge un grand nombre d'options plus spécialisées. Pour obtenir plus de détails, référez-vous à la page `man bltk`.

bltk enregistre les résultats qu'il génère dans un répertoire spécifié dans le fichier de configuration `/etc/bltk.conf` – par défaut, `~/ .bltk/workload.results.number/`. Par exemple, le répertoire `~/ .bltk/reader.results.002/` contient les résultats du troisième test avec la charge de travail **reader** (le premier test n'est pas numéroté). Les résultats sont éparpillés sur plusieurs fichiers texte. Pour convertir ces résultats dans un format facile à lire, exécutez :

```
bltk_report chemin_vers_le_repertoire_des_resultats
```

Les résultats apparaissent maintenant dans un fichier texte nommé **Report** dans le répertoire des résultats. Pour voir les résultats dans un émulateur de terminal, utilisez l'option **-o** :

```
bltk_report -o chemin_vers_le_repertoire_des_resultats
```

2.5. TUNED ET KTUNE

Tuned est un démon qui surveille l'utilisation des composants système et ajuste dynamiquement les paramètres de système en se basant sur les informations de ce suivi. Le paramétrage dynamique permet aux divers composants du système d'être utilisés différemment pendant la durée d'activité de n'importe quel système. Par exemple, le disque dur est beaucoup utilisé lors du démarrage et de la connexion, mais très peu lorsque l'utilisateur travaille avec des applications telles que OpenOffice ou avec de clients email. De la même manière, le CPU et les périphériques réseau sont utilisés de différentes manières à différents moments. **Tuned** surveille l'activité de ces composants et réagit à leurs changements d'utilisation.

Exemple pratique : considérez une station de travail de bureau typique. La plupart du temps, l'interface réseau Ethernet sera inactive. Seuls quelques emails entreront ou sortiront de temps en temps, et/ou

quelques pages web seront chargées. Pour ce type de charges, l'interface réseau n'a pas besoin de s'exécuter à sa vitesse normale tout le temps (ce qui est le cas par défaut). **Tuned** possède un plugin de suivi et paramétrage pour les périphériques réseau, celui-ci peut détecter une activité faible, et automatiquement rabaisser la vitesse de cette interface, résultant ainsi en une plus basse vitesse. Si l'activité de l'interface augmente de manière drastique sur une longue période, par exemple lorsqu'une image DVD est en cours de téléchargement ou lorsqu'un email avec un gros attachement est ouvert, **Tuned** s'en rendra compte et ajustera la vitesse de l'interface au maximum de manière à fournir la meilleure performance possible lorsque l'activité est tant importante. Ce principe est aussi utilisé pour d'autres plugins pour CPU et disques durs.

Les périphériques réseau ne sont pas configurés pour se comporter de cette manière par défaut car les changements de vitesse peuvent mettre quelques secondes à prendre effet, ceux-ci peuvent ainsi avoir un impact direct et visible sur l'expérience utilisateur. Des considérations similaires s'appliquent aux plugins de paramétrage du CPU et du disque dur. Lorsqu'un disque dur est arrêté, celui-ci peut prendre quelques secondes pour se remettre en marche, ce qui résulte en un manque de réactivité du système pendant cette période. L'effet secondaire sur la latence est plus léger pour le plugin du CPU, et même s'il est mesurable, il n'est quasiment pas remarquable par l'utilisateur.

Aux côtés de **tuned**, nous vous offrons aussi **ktune**. **Ktune** fut présenté dans Red Hat Enterprise Linux 5.3 en tant que framework et service pour optimiser la performance d'une machine dans des cas d'utilisation spécifiques. Depuis, **ktune** a été amélioré à un tel point que nous l'utilisons maintenant comme la partie statique de notre framework de paramétrage général. **ktune** est principalement utilisé dans les différents profils prédéfinis dans la [Section 2.5.2, « Tuned-adm »](#).

Installez le paquetage **tuned** et ses scripts **systemtap** associés avec la commande :

```
yum install tuned
```

L'installation du paquetage **tuned** installe aussi un exemple de fichier de configuration sur `/etc/tuned.conf` et active le profil par défaut.

Démarrez **tuned** en exécutant :

```
service tuned start
```

Pour lancer **tuned** à chaque fois que la machine démarre, exécutez :

```
chkconfig tuned on
```

Tuned possède des options supplémentaires que vous pouvez utiliser lorsque vous l'exécutez manuellement. Les options disponibles sont :

-d, --daemon

démarre **Tuned** comme un démon plutôt que sur l'avant-plan.

-c, --conf file

utilise un fichier de configuration avec le nom et chemin spécifié, par exemple, `--conf file=/etc/tuned2.conf`. Le défaut est `/etc/tuned.conf`.

-D, --debug

utilise le plus haut niveau de journalisation.

2.5.1. Le fichier `tuned.conf`

Le fichier `tuned.conf` contient des paramètres de configuration pour `tuned`. Par défaut, il se trouve sur `/etc/tuned.conf`, mais vous pouvez spécifier un nom et un emplacement différent en démarrant `tuned` avec l'option `--conffile`.

Le fichier de configuration doit toujours contenir une section `[main]` qui définit les paramètres généraux pour `tuned`. Le fichier contient ensuite une section pour chaque plugin.

La section `[main]` contient les options suivantes :

`interval`

l'intervalle, en secondes, auquel `tuned` devrait suivre et paramétrer le système. La valeur par défaut est `10`.

`verbose`

spécifie si la sortie doit être verbeuse ou non. La valeur par défaut est `False`.

`logging`

spécifie la priorité minimum des messages à journaliser. Dans un ordre descendant, les valeurs acceptables sont : `critical`, `error`, `warning`, `info`, et `debug`. La valeur par défaut est `info`.

`logging_disable`

spécifie la priorité maximum des messages à journaliser, tout message avec cette priorité ou avec une priorité plus basse ne sera pas journalisé. Dans un ordre descendant, les valeurs acceptables sont : `critical`, `error`, `warning`, `info`, et `debug`. La valeur `notset` désactive cette option.

Chaque plugin possède sa propre section, spécifiée avec le nom du plugin dans des crochets. Par exemple : `[CPUTuning]`. Chaque plugin peut posséder ses propres options, mais les options suivantes s'appliquent à tous les plugins :

`enabled`

spécifie si le plugin est activé ou non. La valeur par défaut est `True`.

`verbose`

spécifie si la sortie doit être verbeuse ou non. Si ce n'est pas défini pour ce plugin, la valeur est héritée de `[main]`.

`logging`

spécifie la priorité minimum des messages à journaliser. Si ce n'est pas défini pour ce plugin, la valeur est héritée de `[main]`.

Exemple de fichier de configuration :

```
[main]
interval=10
pidfile=/var/run/tuned.pid
logging=info
logging_disable=notset
```

```
# Disk monitoring section

[DiskMonitor]
enabled=True
logging=debug

# Disk tuning section

[DiskTuning]
enabled=True
hdparm=False
alpm=False
logging=debug

# Net monitoring section

[NetMonitor]
enabled=True
logging=debug

# Net tuning section

[NetTuning]
enabled=True
logging=debug

# CPU monitoring section

[CPUMonitor]
# Enabled or disable the plugin. Default is True. Any other value
# disables it.
enabled=True

# CPU tuning section

[CPUTuning]
# Enabled or disable the plugin. Default is True. Any other value
# disables it.
enabled=True
```

2.5.2. Tuned-adm

Dans de nombreux cas, l'audit et l'analyse détaillée d'un système peut prendre beaucoup de temps, ceci ne vaut peut être pas les quelques watts que vous économiserez en effectuant ceci. Auparavant, l'unique alternative était d'utiliser les valeurs par défaut. Ainsi, Red Hat Enterprise Linux 6 inclut des profils séparés pour des cas d'utilisation spécifiques, pour offrir une alternative entre ces deux extrêmes, et ce avec l'outil `tuned-adm` qui vous permet de passer facilement d'un profil à l'autre en utilisant la ligne de commande. Red Hat Enterprise Linux 6 inclut un certain nombre de profils prédéterminés pour des cas d'utilisation typiques que vous pouvez simplement sélectionner et activer avec la commande `tuned-adm`, mais vous pouvez aussi créer, modifier, ou supprimer des profils vous-même.

Pour lister tous les profils disponibles et identifier le profil actuellement actif, exécutez :

```
tuned-adm list
```

Pour afficher le profil actuellement actif uniquement, exécutez :

```
tuned-adm active
```

Pour passer à l'un des profils disponibles, exécutez :

```
tuned-adm profile nom_du_profil
```

par exemple :

```
tuned-adm profile server-powersave
```

Pour désactiver le paramétrage :

```
tuned-adm off
```

Lorsque vous installez **tuned** pour la première fois, le profil par **défaut** sera actif. Red Hat Enterprise Linux 6 inclut aussi les profils prédéfinis suivants :

défaut

le profil d'économie d'énergie par défaut. Il a l'impact le plus léger sur les économies d'énergie de tous les profils disponibles et n'active que les CPU et plugins disque de **tuned**.

desktop-powersave

un profil d'économie d'énergie pour systèmes de bureau. Active l'économie d'énergie ALPM pour les adaptateurs d'hôtes SATA (reportez-vous à la [Section 3.6, « Aggressive Link Power Management »](#)) ainsi que les plugins CPU, Ethernet, et disque de **tuned**.

server-powersave

un profil d'économie d'énergie pour systèmes de serveurs. Active l'économie d'énergie ALPM pour les adaptateurs d'hôtes SATA, désactive les interrogations de CD-ROM avec HAL (reportez-vous à la page `man hal-disable-polling`, et active les plugins CPU et disque de **tuned**.

laptop-ac-powersave

un profil d'économie d'énergie d'impact moyen pour ordinateurs portables fonctionnant sur batterie. Active l'économie d'énergie ALPM pour les adaptateurs hôte SATA, WiFi, ainsi que pour les plugins CPU, Ethernet, et disque de **tuned**.

laptop-battery-powersave

un profil d'économie d'énergie d'impact fort pour les ordinateurs portables fonctionnant sur batterie. Il active tous les mécanismes d'économie d'énergie des profils précédents, il active aussi le planificateur d'économie d'énergie multicœurs pour des systèmes avec peu de réveils, vérifie que le gouverneur `ondemand` est bien en marche, et que l'économie d'énergie audio AC97 est bien activée. Vous pouvez utiliser ce profil pour économiser un maximum d'énergie sur n'importe quel type de système, pas seulement sur des ordinateurs portables en mode batterie. Ce profil est activé au compromis d'un impact remarquable sur la performance, plus particulièrement sur la latence de disque et les E/S réseau.

throughput-performance

un profil serveur pour un paramétrage de performance de débit typique. Il désactive les mécanismes d'économie d'énergie **tuned** et **ktune**, active les paramètres **sysctl** qui améliorent la performance de débit de votre disque et des E/S de réseau, et passe au **deadline scheduler**.

latency-performance

un profil serveur pour un paramétrage de performance de la latence typique. Il désactive les mécanismes d'économie d'énergie **tuned** et **ktune**, et active les paramètres **sysctl** qui améliorent la performance de la latence de vos E/S de réseau.

Tous les profils sont stockés dans différents sous-répertoires sous **/etc/tune-profiles**. Ainsi, **/etc/tune-profiles/desktop-powersave** contient tous les fichiers et paramètres nécessaires à ce profil. Chacun de ces répertoires peut contenir jusqu'à quatre fichiers :

tuned.conf

la configuration pour que le service **tuned** soit activé pour ce profil.

sysctl.ktune

les paramètres **sysctl** utilisés par **ktune**. Le format est identique au fichier **/etc/sysconfig/sysctl** (reportez-vous aux pages man **sysctl** et **sysctl.conf**).

ktune.sysconfig

le fichier de configuration de **ktune**, habituellement **/etc/sysconfig/ktune**.

ktune.sh

un script shell de type **init** utilisé par le service **ktune** qui peut exécuter des commandes spécifiques lors du démarrage du système pour paramétrer celui-ci.

La plus simple manière de faire un nouveau profil est d'en copier un existant. Le profil **laptop-battery-powersave** contient un riche ensemble de paramètres prêt à utiliser et constitue ainsi un bon point de départ. Copiez simplement la totalité du répertoire sous un nouveau nom de profil comme suit :

```
cp -a /etc/tune-profiles/laptop-battery-powersave/ /etc/tune-profiles/myprofile
```

Modifiez tout fichier du nouveau profil afin qu'il corresponde à vos besoins personnels. Par exemple, si vous nécessitez la détection de changements CD, vous pourriez désactiver cette optimisation en mettant la ligne correspondante en commentaire dans le script **ktune.sh** :

```
# Disable HAL polling of CDRoms
# for i in /dev/scd*; do hal-disable-polling --device $i; done > /dev/null
2>&1
```

2.6. DEVICEKIT-POWER ET DEVKIT-POWER

Dans Red Hat Enterprise Linux 6 **DeviceKit-power** assume les fonctions de gestion de l'alimentation qui faisaient partie de **HAL** ainsi que certaines fonctions qui faisaient partie du **Gestionnaire de l'alimentation GNOME** (GNOME Power Manager) dans des versions précédentes de Red Hat Enterprise Linux (reportez-vous à la [Section 2.7, « Gestionnaire d'alimentation GNOME »](#)). **DeviceKit-**

power fournit un démon, un API, et un ensemble d'outils de ligne de commande. Chaque source d'alimentation du système est représentée en tant que périphérique, qu'il s'agisse d'un périphérique physique ou non. Par exemple, une batterie d'ordinateur portable et une source d'alimentation secteur sont toutes les deux considérées comme des périphériques.

Vous pouvez accéder aux outils de ligne de commande avec la commande `devkit -power` et les options suivantes :

--enumerate, -e

affiche un chemin d'objet pour chaque périphérique d'alimentation sur le système, par exemple :

```
/org/freedesktop/DeviceKit/power/devices/line_power_AC
/org/freedesktop/UPower/DeviceKit/power/battery_BAT0
```

--dump, -d

affiche les paramètres pour tous les périphériques d'alimentation sur le système.

--wakeups, -w

affiche les réveils CPU sur le système.

--monitor, -m

surveille les modifications apportées aux périphériques d'alimentation sur le sys. Par exemple, la connexion ou déconnexion de la source d'alimentation secteur, ou l'épuisement d'une batterie. Pressez sur **Ctrl+C** pour arrêter la surveillance du système.

--monitor-detail

surveille les modifications apportées aux périphériques d'alimentation sur le système. Par exemple, la connexion ou déconnexion de la source d'alimentation secteur, ou l'épuisement d'une batterie. L'option **--monitor-detail** présente davantage de détails que l'option **--monitor**. Pressez sur **Ctrl+C** pour arrêter la surveillance du système.

--show-info chemin_de_l'objet, -i chemin_de_l'objet

affiche toutes les informations disponibles sur un chemin d'objet particulier. Par exemple, pour obtenir des informations sur une batterie de votre système représentée par le chemin d'objet `/org/freedesktop/UPower/DeviceKit/power/battery_BAT0`, exécutez :

```
devkit-power -i /org/freedesktop/UPower/DeviceKit/power/battery_BAT0
```

2.7. GESTIONNAIRE D'ALIMENTATION GNOME

Le **gestionnaire d'alimentation GNOME** est un démon qui est installé car faisant partie du bureau GNOME. Une grande partie de la fonctionnalité de gestion de l'alimentation que le **gestionnaire d'alimentation GNOME** offrait dans des versions ultérieures de Red Hat Enterprise Linux fait maintenant partie de **DeviceKit-power** dans Red Hat Enterprise Linux 6 (reportez-vous à la [Section 2.6, « DeviceKit-power et devkit-power »](#)). Cependant, le **gestionnaire d'alimentation GNOME** reste la façade de cette fonctionnalité. Au travers d'un applet dans la barre d'état système, le **gestionnaire d'alimentation GNOME** vous avertit des changements d'état de l'alimentation de votre système. Par exemple, lorsque l'alimentation passe de la batterie au secteur. Il rapporte aussi l'état de la batterie, et vous avertit lorsque celle-ci s'épuise.

Le **gestionnaire d'alimentation GNOME** vous permet aussi de configurer certains paramètres de base de gestion de l'alimentation. Pour accéder à ces paramètres, cliquez sur l'icône **gestionnaire d'alimentation GNOME** dans la barre d'état système, puis cliquez sur **Préférences**

L'écran **Préférences du gestionnaire de l'alimentation** contient trois onglets :

- **Alimentation secteur**
- **Alimentation batterie**
- **Général**

Utilisez les onglets **Alimentation secteur** et **Alimentation batterie** afin de spécifier combien de temps doit s'écouler avant que l'affichage ne s'éteigne sur un système inactif, combien de temps doit s'écouler avant qu'un système inactif ne soit mis en veille, et si le système doit interrompre la rotation des disques durs lorsqu'ils ne sont pas utilisés. L'onglet **Alimentation batterie** vous permet aussi de définir la luminosité de l'affichage et de choisir le comportement d'un système avec une batterie extrêmement faible. Par exemple, par défaut, le **gestionnaire d'alimentation GNOME** fait hiberner un système lorsque le niveau de batterie atteint un niveau extrêmement faible. Utilisez l'onglet **Général** pour définir les comportements des boutons (physiques) d'alimentation et de mise en veille de votre système, et pour spécifier les circonstances sous lesquelles l'icône du **gestionnaire d'alimentation GNOME** doit apparaître dans la barre d'état du système.

2.8. AUTRES MANIÈRES D'AUDITER

Red Hat Enterprise Linux 6 offre un certain nombre d'outils avec lesquels procéder à des audits et analyses de systèmes. La plupart de ceux-ci peuvent être utilisés comme sources d'informations supplémentaires au cas où vous souhaiteriez vérifier ce que vous avez déjà découvert ou au cas où vous auriez besoin d'informations plus approfondies sur certaines parties. La majorité de ces outils sont aussi utilisés pour le paramétrage de performance. Ils incluent :

vmstat

vmstat vous donne des détails sur les processus, la mémoire, la pagination, les E/S de bloc, les dérivements, et l'activité du CPU. Utilisez cette commande pour voir de plus près ce que le système fait et où il est occupé.

iostat

iostat est similaire à **vmstat**, mais seulement pour les E/S des périphériques de bloc. Cette commande fournit aussi une sortie et des statistiques plus verbeuses.

blktrace

blktrace est un programme de suivi détaillé des E/S de bloc. Il décompose les informations en blocs uniques associés à des applications. Il se révèle très utile en conjonction avec **diskdevstat**.

CHAPITRE 3. INFRASTRUCTURE ET MÉCANIQUE PRINCIPALE

3.1. ÉTATS INACTIFS DU CPU

Les CPUs avec architecture x86 prennent en charge différents états dans lesquels certaines parties du CPU sont désactivées ou s'exécutent sur des paramètres de performance plus bas. Ces états, connus sous le nom de *C-states*, permettent aux systèmes d'économiser de l'énergie en désactivant partiellement les CPUs qui ne sont pas utilisés. Les *C-states* sont numérotés de manière ascendante, commençant par C0, et les numéros plus élevés représentent une fonctionnalité du CPU réduite et une économie d'énergie plus importante. Les *C-states* d'un numéro donné sont relativement similaires entre différents processeurs, même si les détails exacts des fonctionnalités spécifiques de l'état peut varier en fonction de la famille de processeurs. Les *C-states* de 0 à 3 sont définis comme suit :

C0

l'état d'opération, ou d'exécution. Dans cet état, le CPU fonctionne et n'est pas inactif.

C1, Halt

un état dans lequel le processeur n'exécute pas d'instructions mais n'est habituellement pas dans un état de basse alimentation. Le CPU peut continuer ses calculs avec pratiquement aucun délai. Tous les processeurs offrant des états *C-States* doivent pouvoir prendre cet état en charge. Les processeurs Pentium 4 prennent en charge un état C1 amélioré nommé C1E qui est en fait un état de consommation d'électricité moindre.

C2, Stop-Clock

un état où l'horloge est gelée pour le processeur, mais qui reste dans son état complet pour ses registres et caches, de manière à ce que les processeurs recommencent leurs calculs dès le démarrage de l'horloge. Il s'agit d'un état optionnel.

C3, Sleep

un état où le processeur entre réellement dans un état de sommeil et n'a pas besoin de garder son cache à jour. À cause de ceci, le réveiller de cet état prendra considérablement plus de temps que le réveiller de C2. Une fois de plus, il s'agit ici d'un état optionnel.

Les CPUs d'Intel récents avec la microarchitecture "Nehalem" proposent un nouveau *C-state*, C6, permettant de réduire le voltage du CPU à zéro, mais réduit aussi la consommation d'énergie de 80% à 90%. Le noyau dans Red Hat Enterprise Linux 6 inclut des optimisations pour ce nouvel état *C-state*.

3.2. UTILISATION DES GOUVERNEURS `CPUFREQ`

L'une des manières les plus efficaces de réduire la consommation d'énergie et la sortie de chaleur sur votre système est d'utiliser `CPUFreq`. `CPUFreq` – aussi connu sous le nom de mise à l'échelle de la vitesse du CPU – permet d'ajuster la vitesse d'horloge du processeur spontanément. Ceci permet au système de fonctionner à une vitesse d'horloge réduite afin d'économiser de l'énergie. Les règles pour les changements de fréquences, que ce soit vers une vitesse d'horloge plus ou moins élevée, ou à quel moment les changer, sont définies par le gouverneur `CPUFreq`.

Le gouverneur définit les caractéristiques du CPU du système, qui elles-mêmes affecteront la performance du CPU. Chaque gouverneur possède son propre comportement, but, et sa propre pertinence en termes de charge de travail. Cette section décrit comment choisir et configurer un gouverneur `CPUFreq`, les caractéristiques de chaque gouverneur, et le type de charge de travail que chaque gouverneur peut effectuer.

Les principaux problèmes autour de la gestion de l'alimentation sont :

- Réduction de chaleur pour les serveurs
- Augmentation de l'autonomie de la batterie pour les ordinateurs portables

En tant que règle, la baisse de consommation d'énergie d'un composant spécifique (ou du système en tant que tout) conduira à une baisse de chaleur et naturellement, de performance. Ainsi, vous devriez étudier et tester minutieusement les baisses de performance occasionnées par les configurations que vous faites, particulièrement pour les systèmes de missions critiques.

Les sections suivantes expliquent en quoi une performance optimale du matériel peut bénéficier à votre système en termes de consommation d'énergie.

3.2.1. Types de gouverneurs CPUfreq

Cette section liste et décrit les différents types de gouverneurs CPUfreq disponibles dans Red Hat Enterprise Linux 6.

cpufreq_performance

Le gouverneur Performance force le CPU à utiliser la plus haute fréquence d'horloge possible. Cette fréquence est statiquement réglée, et ne changera pas. Ainsi, ce gouverneur en particulier n'offre *aucune économie d'énergie*. Il n'est utile que pour de longues heures de charges de travail intensives, et lorsque le CPU est rarement (ou n'est pas) inactif.

cpufreq_powersave

Parcontre, le gouverneur Powersave force le CPU à utiliser la plus basse fréquence d'horloge possible. Cette fréquence est réglée statiquement, et ne changera pas. Ainsi, ce gouverneur en particulier offre une économie d'énergie maximale, mais ceci au prix d'une *performance de CPU plus basse*.

Le terme "powersave" peut parfois prêter à confusion, puisqu'un CPU lent (en principe) sur une charge pleine consomme plus d'énergie qu'un processeur rapide qui n'est pas chargé. Ainsi, alors qu'il est recommandé de paramétrer le CPU pour qu'il utilise le gouverneur Powersave lors de périodes d'activité restreintes, toute charge de travail lourde à ce moment peut causer au système de consommer plus d'énergie.

Le gouverneur Powersave est, en d'autres termes, plus un « limiteur de vitesse » qu'un « économiseur d'énergie ». Il est particulièrement utile pour des systèmes et environnements dans lesquels une surchauffe pourrait être un problème.

cpufreq_ondemand

Le gouverneur Ondemand est un gouverneur dynamique qui permet au CPU d'atteindre une fréquence d'horloge maximale lorsque la charge du système est importante, et d'atteindre une fréquence minimale lorsque le système est inactif. Tandis que ceci permet d'ajuster la consommation d'énergie en fonction de la charge du système, cette opération se produit au détriment de la *latence entre les changements de fréquences*. Ainsi, la latence peut contrebalancer les bénéfices de performance/d'énergie du gouverneur Ondemand si le système change trop souvent entre inactivité et charges de travail intenses.

Pour la plupart des systèmes, le gouverneur Ondemand peut fournir le meilleur compromis qui soit entre émissions de chaleur, consommation d'énergie, performance, et manageabilité. Lorsque le système est uniquement occupé à certains moments spécifique de la journée, le gouverneur Ondemand changera automatiquement entre les fréquences maximale et minimale en fonction de la charge et sans la moindre autre intervention.

cpufreq_userspace

Le gouverneur Userspace permet aux programmes de l'espace utilisateur (ou à tout autre processus exécuté en tant que root) de définir la fréquence. Ce gouverneur est normalement utilisé avec le démon `cpuspeed`. De tous les gouverneurs, Userspace est le plus personnalisable ; et selon sa configuration, il peut fournir le meilleur équilibre entre performance et consommation pour votre système.

cpufreq_conservative

Tout comme le gouverneur Ondemand, le gouverneur Conservative ajuste aussi la fréquence de l'horloge en fonction de l'utilisation (comme le gouverneur Ondemand). Cependant, tandis que gouverneur le fait d'une manière plus agressive (du maximum au minimum, et vice-versa), le gouverneur Conservative change de fréquence de manière plus graduelle.

Ceci signifie que plutôt que de simplement choisir entre maximum et minimum, le gouverneur Conservative s'ajustera à une fréquence d'horloge qu'il reconnaîtra appropriée en fonction de la charge. Alors que ceci peut offrir d'importantes économies d'énergie, cette opération est effectuée aux dépend d'une *latence bien plus élevée* que celle du gouverneur Ondemand.



NOTE

Vous pouvez activer un gouverneur à l'aide des tâches `cron`. Ceci vous permettra de définir automatiquement des gouverneurs particuliers à des moments spécifique de la journée. Ainsi, vous pouvez spécifier un gouverneur de basse latence pendant les périodes d'inactivité (par exemple hors des heures de travail traditionnelles), puis retourner à un gouverneur de fréquence plus importante durant les heures de charges de travail plus importantes.

Pour des instructions sur comment activer un gouverneur spécifique, veuillez vous reporter à la [Procédure 3.2, « Activer un gouverneur CPUfreq »](#) dans la [Section 3.2.2, « Installation de CPUfreq »](#).

3.2.2. Installation de CPUfreq

Avant de sélectionner et de configurer un gouverneur CPUfreq, vous devrez commencer par acquérir le pilote CPUfreq approprié.

Procédure 3.1. Comment ajouter un pilote CPUfreq

1. Utilisez la commande suivante pour voir quels pilotes CPUfreq sont disponibles pour votre système :

```
ls /lib/modules/[version du
noyau]/kernel/arch/[architecture]/kernel/cpu/cpufreq/
```

2. Utilisez `modprobe` afin d'ajouter le pilote CPUfreq approprié.

```
modprobe [pilote CPUfreq]
```

Lors de l'utilisation de la commande ci-dessus, assurez-vous de bien supprimer le suffixe de nom de fichier `.ko`.

IMPORTANT

Lors de la sélection d'un pilote CPUfreq approprié, choisissez toujours `acpi-cpufreq` plutôt que `p4-clockmod`. Alors que l'utilisation du pilote `p4-clockmod` réduit la fréquence d'horloge d'un CPU, il ne réduit pas son voltage. D'un autre côté, `acpi-cpufreq`, réduit le voltage ainsi que la fréquence d'horloge du CPU, permettant une moindre consommation d'énergie et sortie de chaleur pour chaque réduction d'unité en performance.

- Une fois que le pilote CPUfreq est installé, vous pourrez voir quel gouverneur le système utilise avec :

```
cat /sys/devices/system/cpu/cpu0/cpufreq/scaling_governor
```

Vous pourrez aussi voir quels gouverneurs sont disponibles pour un CPU spécifique avec :

```
cat /sys/devices/system/cpu/[ID du
cpu]/cpufreq/scaling_available_governors
```

Certains gouverneurs CPUfreq peuvent ne pas être à votre disposition. Dans ce cas, utilisez `modprobe` pour ajouter les modules de noyau nécessaires pour activer le gouverneur CPUfreq que vous souhaitez utiliser en particulier. Ces modules de noyau sont disponibles dans `/lib/modules/[version du noyau]/kernel/drivers/cpufreq/`.

Procédure 3.2. Activer un gouverneur CPUfreq

- Si un gouverneur spécifique n'est pas disponible pour votre CPU, utilisez `modprobe` pour activer le gouverneur que vous souhaitez utiliser. Par exemple, si le gouverneur `ondemand` n'est pas disponible pour votre CPU, utilisez la commande suivante :

```
modprobe cpufreq_ondemand
```

- Une fois qu'un gouverneur est listé comme disponible pour votre CPU, vous pourrez l'activer à l'aide de :

```
echo [gouverneur] >
/sys/devices/system/cpu/cpu0/cpufreq/scaling_governor
```

3.2.3. Paramétrer la politique CPUfreq et la vitesse

Une fois que vous aurez choisi un gouverneur CPUfreq approprié, vous pourrez paramétrer la vitesse de chaque CPU à l'aide des paramètres se trouvant dans `/sys/devices/system/cpu/[ID du cpu]/cpufreq/`. Ces paramètres sont :

- `cpuinfo_min_freq` – Affiche la fréquence minimale de fonctionnement disponible du CPU (en KHz).
- `cpuinfo_max_freq` – Affiche la fréquence maximale de fonctionnement disponible du CPU (en KHz).
- `scaling_driver` – Affiche quel pilote de CPUfreq est utilisé pour paramétrer la fréquence sur ce CPU.

- **scaling_available_governors** – Affiche les gouverneurs CPUfreq disponibles dans ce noyau. Si vous souhaitez utiliser un gouverneur CPUfreq qui n'est pas listé dans ce fichier, reportez-vous à la [Procédure 3.2, « Activer un gouverneur CPUfreq »](#) in [Section 3.2.2, « Installation de CPUfreq »](#) pour obtenir des instructions.
- **scaling_governor** – Affiche quel gouverneur CPUfreq est actuellement en cours d'utilisation. Pour utiliser un différent gouverneur, utilisez simplement `echo [gouverneur] > /sys/devices/system/cpu/[ID du cpu]/cpufreq/scaling_governor` (reportez-vous à la [Procédure 3.2, « Activer un gouverneur CPUfreq »](#) dans la [Section 3.2.2, « Installation de CPUfreq »](#) pour obtenir plus d'informations).
- **cpufreq_info** – Affiche la vitesse actuelle du CPU (en KHz).
- **scaling_available_frequencies** – Liste les fréquences disponible pour le CPU, en KHz.
- **scaling_min_freq** et **scaling_max_freq** – Définissent les *limites de la politique* du CPU, en KHz.



IMPORTANT

Lors de la définition des limites de la politique, vous devriez définir **scaling_max_freq** avant **scaling_min_freq**.

- **affected_cpus** – Liste les CPUs qui requièrent un logiciel de coordination de fréquence.
- **scaling_setspeed** – Utilisé pour modifier la vitesse d'horloge du processeur, en KHz. Vous pouvez uniquement définir une vitesse dans les limites imposées par la politique du processeur (c'est-à-dire par **scaling_min_freq** et **scaling_max_freq**).

Pour voir la valeur actuelle de chaque paramètre, utilisez `cat [paramètre]`. Par exemple, pour voir la vitesse actuelle de `cpu0` (en KHz), utilisez :

```
cat /sys/devices/system/cpu/cpu0/cpufreq/cpuinfo_cur_freq.
```

Pour modifier la valeur de chaque paramètre, utilisez `echo [valeur] > /sys/devices/system/cpu/[ID du cpu]/cpufreq/[paramètre]`. Par exemple, pour régler la vitesse d'horloge de processeur minimum de `cpu0` sur 360 KHz, utilisez :

```
echo 360000 > /sys/devices/system/cpu/cpu0/cpufreq/scaling_min_freq
```

3.3. SUSPENSION ET REPRISE

Lorsqu'un système est suspendu, le noyau appelle les pilotes à stocker leurs états puis les décharge. Lorsque le système est repris, il recharge ces pilotes, qui tentent alors de reprogrammer leurs périphériques. La capacité des pilotes à accomplir cette tâche déterminera si le système peut être repris avec succès.

Les pilotes vidéo sont particulièrement problématiques car la spécification *ACPI* (Advanced Configuration and Power Interface) ne requiert pas du microprogramme du système qu'il ne reprogramme le matériel vidéo. Ainsi, à moins que les pilotes vidéo ne soient capables de programmer le matériel depuis un état non-initialisé, ils pourront empêcher la reprise du système.

Red Hat Enterprise Linux 6 inclut une meilleure prise en charge des chipsets graphiques, ce qui permet à la suspension et à la reprise de fonctionner sur un plus grand nombre de plateformes. La prise en

charge des chipsets NVIDIA a été particulièrement améliorée ; notamment pour les séries GeForce 8800.

3.4. NOYAU TICKLESS

Auparavant, le noyau Linux interrompait périodiquement chaque CPU du système à une fréquence prédéterminée – 100 Hz, 250 Hz, ou 1000 Hz, selon la plateforme. Le noyau questionnait le CPU sur les processus qu'il exécutait, puis utilisait les résultats pour compter les processus et équilibrer les charges. Connue sous le nom de *tic d'horloge*, le noyau effectuait ces interruptions peu importe l'état d'alimentation du CPU. Ainsi, même un CPU inactif devait répondre à jusqu'à 1000 requêtes de la sorte par seconde. Sur les systèmes qui implémentaient des mesures d'économies d'énergie pour les CPU inactifs, le tic d'horloge empêchait le CPU de rester inactif assez longtemps pour pouvoir bénéficier de ces économies d'énergie.

Le noyau dans Red Hat Enterprise Linux 6 exécute *tickless* : qui remplace les anciennes interruptions de l'horloge périodique par des interruptions à la demande. Ainsi, les CPU inactifs peuvent rester inactifs jusqu'à ce qu'une nouvelle tâche soit placée en file d'attente pour traitement, et les CPU se trouvant dans des états de basse alimentation peuvent rester dans ces états plus longtemps.

3.5. GESTION DE L'ALIMENTATION EN ÉTAT ACTIF

La *gestion de l'alimentation en état actif* (ASPM, de l'anglais Active-State Power Management) permet de réaliser des économies d'énergie dans le sous-système *interconnexion de composants périphériques* (PCI Express, ou PCIe) en définissant un état d'alimentation plus faible pour les liens PCIe lorsque les périphériques auxquels ils sont connectés ne sont pas en cours d'utilisation. ASPM contrôle l'état de l'alimentation des deux côtés du lien, et permet d'économiser de l'énergie dans le lien même lorsque le périphérique se trouvant à la fin de celui-ci est dans un état d'alimentation complète.

Lorsque ASPM est activé, la latence de périphérique augmente à cause du temps requis pour effectuer la transition du lien entre les différents états d'alimentation. ASPM possède trois politiques afin de déterminer les états d'alimentation :

default

établit les états d'alimentation du lien PCIe en fonction des valeurs par défaut spécifiées par le microprogramme du système (par exemple, BIOS). Ceci est l'état par défaut d'ASPM.

powersave

règle ASPM de manière à économiser de l'énergie à chaque occasion, sans tenir compte de la performance en résultant.

performance

désactive ASPM afin de permettre aux liens PCIe d'opérer avec une performance maximale.

Les politiques ASPM sont définies dans `/sys/module/pci_esp/parameters/policy`, mais elles peuvent aussi être spécifiées lors du démarrage avec le paramètre de noyau `pci_esp`, où `pci_esp=off` désactive ASPM et `pci_esp=force` active ASPM, même sur les périphériques ne prenant pas ASPM en charge.



AVERTISSEMENT

Si `pcie_aspm=force` est défini, le matériel qui ne prend pas en charge ASPM peut provoquer une cessation de réponse du système. Avant de définir `pcie_aspm=force`, assurez-vous que tout le matériel PCIe sur le système prend ASPM en charge.

3.6. AGGRESSIVE LINK POWER MANAGEMENT

ALPM (de l'anglais, Aggressive Link Power Management) est une technique d'économie d'énergie qui aide le disque à économiser l'énergie en paramétrant un lien SATA sur une faible puissance lors d'un temps d'inactivité (lorsqu'il n'y a pas d'E/S). ALPM règle automatiquement le lien SATA sur un état actif une fois que les requêtes d'E/S sont en file d'attente sur ce lien.

Les économies d'énergie offertes par ALPM se font aux dépens de la latence de disque. Ainsi, ALPM ne devrait être utilisé que dans le cas où le système connaîtra de longues périodes d'inactivité d'E/S.

ALPM est uniquement disponible sur les contrôleurs SATA qui utilisent les interfaces AHCI (de l'anglais, Advanced Host Controller Interface). Pour obtenir plus d'informations sur AHCI, reportez-vous à <http://www.intel.com/technology/serialata/ahci.htm>.

Lorsque disponible, ALPM est activé par défaut. ALPM possède trois niveaux :

min_power

Ce mode définit le lien sur son état le plus bas (SLUMBER) lorsqu'il n'y a pas d'E/S sur le disque. Ce mode est utile lorsque de longues périodes d'inactivité sont à prévoir.

medium_power

Ce mode définit le lien sur le second état le plus bas (PARTIAL) lorsqu'il n'y a pas d'E/S sur le disque. Ce mode est conçu afin de permettre des transitions entre états d'alimentation (par exemple lorsqu'il y a des alternances entre une utilisation intense d'E/S et une période d'inactivité d'E/S) tout en limitant au maximum l'impact sur la performance.

Le mode `medium_power` permet au lien de faire la transition entre les états PARTIAL et alimentation complète (c'est-à-dire "ACTIVE"), et ce en fonction de la charge. Remarquez qu'il n'est pas possible d'établir une transition de lien directement de PARTIAL à SLUMBER (et vice-versa) ; dans ce cas, les états d'alimentation ne peuvent pas passer de l'un à l'autre sans transiter par l'état ACTIVE au préalable.

max_performance

ALPM est désactivé, le lien ne peut pas entrer dans un état de faible alimentation lorsqu'il n'y a pas d'E/S sur le disque.

Pour vérifier si vos adaptateurs hôtes SATA prennent en charge ALPM, vous pouvez vérifier que le fichier `/sys/class/scsi_host/host*/link_power_management_policy` existe. Pour changer les paramètres, écrivez simplement les valeurs décrites dans cette section sur ces fichiers, ou affichez les fichiers pour vérifier le paramètre actuel.



IMPORTANT

Définir ALPM sur `min_power` ou `medium_power` désactivera automatiquement la fonctionnalité "Hot Plug".

3.7. OPTIMISATION D'ACCÈS AU DISQUE RELATIME

Le standard POSIX requiert que les systèmes d'exploitation effectue la maintenance des méta-données du système de fichiers qui enregistrent la dernière fois que chaque fichier a été accédé. Cette date de dernière modification est appelée `atime`, et la maintenir requiert une constante série d'opérations d'écriture sur le stockage. Ces écritures gardent les périphériques de stockage et leurs liens occupés et allumés. Comme peu d'applications utilisent les données `atime`, l'activité de ce périphérique de stockage gaspille de l'énergie. De plus, l'écriture sur le stockage se produirait même si le fichier n'était pas lu à partir du stockage, mais à partir du cache. Depuis quelques temps, le noyau Linux prend en charge une option `noatime` pour `mount` et n'écrit pas de données `atime` sur les systèmes de fichiers montés avec cette option. Cependant, l'arrêt de cette fonctionnalité est problématique car certaines applications reposent sur les données `atime`, et elles échoueront si ces données ne sont pas disponibles.

Le noyau utilisé dans Red Hat Enterprise Linux 6 prend en charge une autre alternative – `relatime`. `Relatime` effectue la maintenance de `atime`, mais pas à chaque fois qu'on accède à un fichier. Avec l'activation de cette option, les données `atime` sont écrites sur le disque seulement si le fichier a été modifié depuis que la dernière fois que les données `atime` ont été mises à jour (`mtime`), ou si le dernier accès au fichier a eu lieu depuis plus d'un certain temps (par défaut, un jour).

Par défaut, tous les systèmes de fichiers sont maintenant montés avec `relatime` activé. Pour supprimer cette fonctionnalité sur un système entier, utilisez le paramètre de démarrage `default_relatime=0`. Si `relatime` est activé sur un système par défaut, vous pouvez le supprimer sur n'importe quel système de fichiers en particulier en montant ce système de fichiers avec l'option `norelatime`. Finalement, pour changer le délai à partir duquel le système devra mettre à jour les données `atime` d'un fichier, utilisez le paramètre `relatime_interval=`, en spécifiant la période en secondes. La valeur par défaut est `86400`.

3.8. LIMITATION DE LA CONSOMMATION

Red Hat Enterprise Linux 6 prend en charge les fonctionnalités de limitation de la consommation trouvées dans le matériel récent, tel que dans le `DPC` (Dynamic Power Capping, ou limitation de consommation dynamique) de HP et la technologie de gestionnaire de noeuds (Node Manager) d'Intel. La limitation de consommation permet aux administrateurs de limiter la consommation d'énergie des serveurs, mais elle permet aussi aux gérants de planifier les centres de données de manière plus efficace, car le risque de surcharger les réserves d'énergie sont grandement réduits. Les gérants peuvent ainsi placer plusieurs serveurs sur le même espace physique et savoir que si la consommation est limitée, la demande d'énergie lors de lourdes charges n'excédera pas la quantité disponible.

Dynamic Power Capping de HP

Le Dynamic Power Capping est une fonctionnalité disponible sur les serveurs sélectionnés ProLiant et BladeSystem, elle permet aux administrateurs systèmes de limiter la consommation de l'alimentation d'un serveur ou d'un groupe de serveurs. Cette limitation est un limite définitive que le serveur ne pourra pas dépasser, peu importe sa charge de travail. Celle-ci n'aura aucun effet jusqu'à ce que le serveur n'atteigne la limite de consommation. À ce moment, un processeur de gestion ajustera les états P (ou P-states) du CPU, et la limitation d'horloge afin de limiter l'énergie consommée.

Le Dynamic Power Capping modifie le comportement du CPU indépendamment du système d'exploitation, cependant le microprogramme de HP `iLO2` (integrated Lights-Out 2) permet aux

systèmes d'exploitation d'accéder au processeur de gestion et les application de l'espace utilisateur peuvent ainsi questionner le processeur de gestion. Le noyau utilisé dans Red Hat Enterprise Linux 6 inclut un pilote pour les microprogrammes de HP iLO et iLO2, qui permettent de questionner les processeurs de gestion sur `/dev/hpilo/dXccbM`. Le noyau inclut aussi une extension de l'interface `hwmon sysfs` pour prendre en charge les fonctionnalités de limitation de la consommation, ainsi qu'un pilote `hwmon` pour mesureurs d'alimentation ACPI 4.0 qui utilisent l'interface `sysfs`. Ensemble, ces fonctionnalités permettent au système d'exploitation et aux outils de l'espace utilisateur de lire la valeur configurée pour la limitation de consommation, ainsi que la consommation actuelle du système.

Pour obtenir plus de détails sur la limitation de consommation et le Dynamic Power Capping de HP, reportez-vous à *HP Power Capping and HP Dynamic Power Capping for ProLiant Servers*, disponible sur <http://h20000.www2.hp.com/bc/docs/support/SupportManual/c01549455/c01549455.pdf>.

Gestionnaire de noeuds Intel

Le gestionnaire de noeuds Intel impose une limitation de la consommation aux systèmes, à l'aide du processeur des états P et T (T-states et P-states) pour limiter la performance du CPU et par conséquent la consommation de l'alimentation. En définissant une politique de gestion de l'alimentation, les administrateurs peuvent configurer les systèmes afin qu'ils utilisent moins d'énergie lorsque les charges du système sont basses, par exemple pendant la nuit ou les week-ends.

Le gestionnaire de noeuds Intel ajuste la performance du CPU à l'aide de l'*OSPM* (de l'anglais, Operating System-directed configuration and Power Management) avec l'*ACPI* (Advanced Configuration and Power Interface) standard. Lorsque le gestionnaire de noeuds Intel notifie le pilote *OSPM* des changements des T-states, le pilote procède aux changements correspondants sur le processeur des P-states. De la même manière, lorsque le gestionnaire de noeuds Intel notifie le pilote *OSPM* des changements des P-states, celui-ci effectuera les changements correspondants sur les T-states. Ces changements s'effectuent automatiquement et ne requièrent pas l'ingérence du système d'exploitation. Les administrateurs configurent et surveillent le gestionnaire de noeuds Intel avec un logiciel *DCM* (gestionnaire de centre de données Intel, de l'anglais, Intel Data Center Manager).

Pour obtenir plus de détails sur le gestionnaire de noeuds Intel, reportez-vous à *Node Manager – A Dynamic Approach To Managing Power In The Data Center*, disponible sur <http://communities.intel.com/docs/DOC-4766>

3.9. GESTION AMÉLIORÉE DE L'ALIMENTATION DES GRAPHISMES

Red Hat Enterprise Linux 6 permet d'économiser de l'énergie sur les graphismes et moniteurs en éliminant plusieurs sources de consommation non-nécessaires.

Regénération d'horloge LVDS

La *signalisation différentielle à basse tension* (de l'anglais, low-voltage differential signalling, ou LVDS) est un système transportant des signaux électroniques sur des fils de cuivre. Une application importante de ce système est la transmission d'informations de pixels sur les écrans *LCD* (de l'anglais, liquid crystal display) d'ordinateurs portables. Tous les affichages possèdent des *fréquences de rafraîchissement* – la fréquence à laquelle ils reçoivent des nouvelles données depuis un contrôleur graphique et redessinent l'image à l'écran. Habituellement, l'écran reçoit de nouvelles données soixante fois par seconde (une fréquence de 60 Hz). Lorsqu'un écran et un contrôleur graphique sont liés par LVDS, le système LVDS utilise de l'énergie à chaque cycle de rafraîchissement. En période d'inactivité, la fréquence de rafraîchissement de nombreux écrans LCD peut descendre jusqu'à 30 Hz sans que cela ne puisse particulièrement se remarquer (contrairement aux moniteurs à *tube cathodique* (ou CRT), pour lesquels la baisse de la fréquence de rafraîchissement produit un scintillement distinct). Le pilote des adaptateurs graphiques Intel, qui est intégré dans le noyau utilisé dans Red Hat Enterprise Linux 6 effectue cette *baisse d'horloge* automatiquement, et permet d'économiser environ 0.5 W lorsque l'écran est inactif.

Activation du rafraîchissement automatique de la mémoire

La *mémoire vive dynamique synchrone* (ou SDRAM) – comme celle utilisée pour la mémoire vidéo dans les adaptateurs graphiques – est rechargée des milliers de fois par seconde de manière à ce que les cellules de mémoire individuelles puissent retenir les données stockées en elles. À part sa fonction principale de gestion de données circulant vers et hors de la mémoire, le contrôleur de mémoire est habituellement responsable de l'initiation de ces cycles de rafraîchissement. Toutefois, SDRAM possède aussi un mode de *rafraîchissement* de basse consommation. Dans ce mode, la mémoire utilise un compteur interne pour générer ses propres cycles de rafraîchissement, permettant ainsi au système de fermer le contrôleur de mémoire sans pour mettre en danger les données présentes en mémoire. Le noyau utilisé dans Red Hat Enterprise Linux 6 peut ainsi déclencher le rafraîchissement automatique de la mémoire avec les adaptateurs graphiques d'Intel lorsqu'ils sont inactifs, ce qui permet économiser autour de 0,8 W.

Réduction de l'horloge GPU

Les GPUs (de l'anglais, graphical processing units) typiques contiennent des horloges internes gouvernant différentes parties de leur circuiterie interne. Le noyau utilisé dans Red Hat Enterprise Linux 6 peut réduire la fréquence de certaines horloges internes des GPUs d'Intel et d'ATI. La réduction du nombre de cycles effectués par les composants GPU sur un temps donné permet d'économiser de l'énergie qu'ils auraient consommé lors de cycles qu'ils n'avaient pas à effectuer. Le noyau réduit automatiquement la vitesse de ces horloges lorsque le GPU est inactif, et l'augmente lorsque l'activité de celui-ci augmente. La réduction des cycles d'horloge des GPUs peut vous économiser jusqu'à 5 W.

Mise hors-tension GPU

Les pilotes graphiques Intel et ATI dans Red Hat Enterprise Linux 6 peuvent détecter lorsqu'aucun moniteur n'est attaché à un adaptateur et éteint ainsi complètement le GPU. Cette fonctionnalité est particulièrement significative pour les serveurs ne possédant pas de moniteur de manière régulière.

3.10. RFKILL

De nombreux systèmes informatiques contiennent des transmetteurs radio, notamment des périphériques Wi-Fi, Bluetooth, et 3G. Ces périphériques consomment de l'électricité, qui est gaspillée lorsque le périphérique n'est pas en cours d'utilisation.

RFKill est un sous-système dans le noyau Linux qui fournit une interface au travers de laquelle les transmetteurs radio d'un système informatique peuvent être questionnés, activés, et désactivés. Lorsque les transmetteurs sont désactivés, ils peuvent être placés dans un état où le logiciel peut les réactiver (un *soft block*), ou dans un état où le logiciel ne peut pas les réactiver (un *hard block*).

Le cœur RFKill fournit l'interface de programmation (API) pour le sous-système. Les pilotes de noyau désignés pour prendre RFKill en charge utilisent cet API pour s'enregistrer auprès du noyau, et incluent des méthodes pour activer et désactiver le périphérique. De plus, le cœur RFKill offre des notifications que les applications de l'utilisateur peuvent interpréter, ainsi que des manières pour que les applications de l'utilisateur puissent questionner les états des transmetteurs.

L'interface RFKill se trouve dans `/dev/rfkill`, qui contient l'état actuel de tous les transmetteurs radio sur le système. Chaque périphérique possède son propre état RFKill enregistré dans `sysfs`. De plus, RFKill fournit des *uevents* pour chaque changement d'état dans un périphérique pour lequel RFKill est activé.

`Rfkill` est un outil de la ligne de commande qui vous permet de questionner et de modifier des périphériques du système qui possèdent RFKill activé. Pour obtenir cet outil, installez le paquet `rfkill`.

Utilisez la commande `rfkill list` pour obtenir une liste des périphériques, chacun d'entre eux aura un *numéro d'index* qui lui est associé, commençant par `0`. Vous pourrez utiliser ce numéro d'index pour dire à `rfkill` de bloquer ou débloquer un service, par exemple :

```
rfkill block 0
```

bloque le premier périphérique RFKill sur le système.

Vous pouvez aussi utiliser `rfkill` pour bloquer certaines catégories de périphériques, ou même tous les périphériques RFKill. Par exemple :

```
rfkill block wifi
```

bloque tous les périphériques Wi-Fi sur le système. Pour bloquer tous les périphériques RFKill, exécutez :

```
rfkill block all
```

Pour débloquent des périphériques, exécutez `rfkill unblock` au lieu de `rfkill block`. Pour obtenir la liste complète des catégories de périphériques que `rfkill` peut bloquer, exécutez `rfkill help`

3.11. OPTIMISATIONS DANS L'ESPACE UTILISATEUR

La réduction de la charge de travail effectuée par le matériel d'un système est fondamentale à l'économie d'énergie. Ainsi, même si les changements décrits dans le [Chapitre 3, Infrastructure et mécanique principale](#) permettent au système d'opérer dans divers états de consommation d'énergie réduite, les applications dans l'espace utilisateur qui requièrent un travail non-nécessaire de la part du matériel du système empêchent le matériel d'entrer dans ces états. Lors du développement de Red Hat Enterprise Linux 6, des audits ont été entrepris dans les domaines suivants afin de réduire les demandes non-nécessaires sur le matériel :

Réveils réduits

Red Hat Enterprise Linux 6 utilise un *noyau tickless* (ou sans « tic », reportez-vous à la [Section 3.4, « Noyau Tickless »](#)), qui permet aux CPUs de rester plus longtemps dans des états inactifs plus profonds. Cependant, le *timer tick* (ou compteur de « tics ») n'est pas la seule source de réveils CPU excessifs, et les appels de fonctions des applications peuvent aussi empêcher les CPUs d'entrer ou de rester en état inactif. Les appels de fonctions non-nécessaires ont été réduits dans plus de 50 applications.

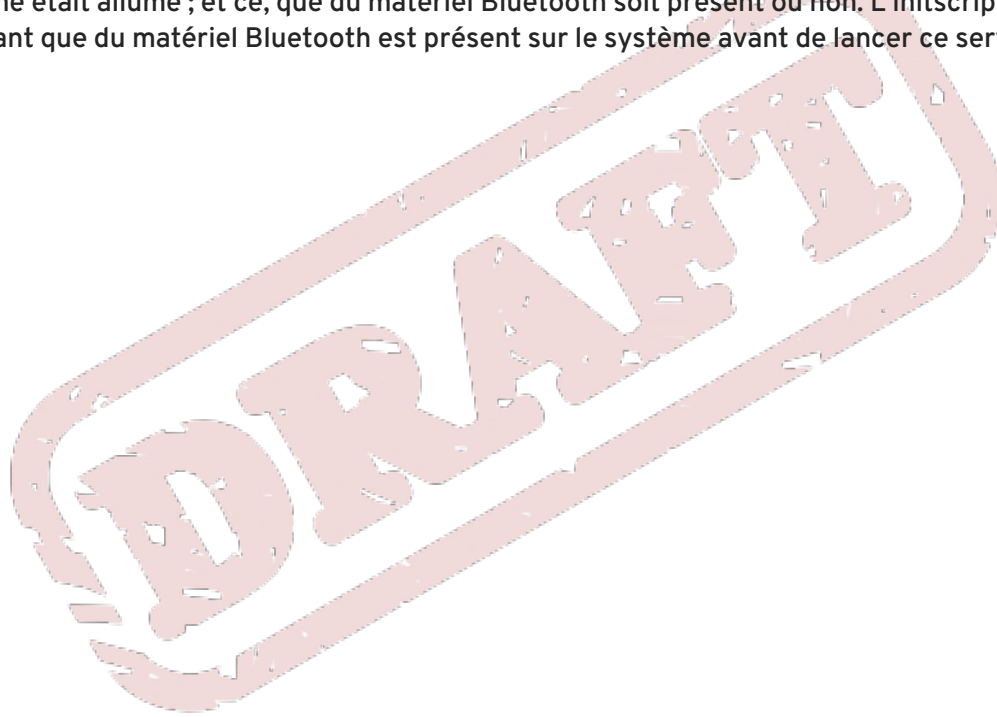
Stockage réduit et E/S réseau

Les entrées ou sorties (E/S) de périphériques de stockage et d'interfaces réseau forcent les périphériques à consommer de l'électricité. Pour les périphériques de stockages et réseau qui proposent des états d'alimentation réduits lorsqu'ils sont inactifs (par exemple, ALPM ou ASPM), ce trafic peut empêcher le périphérique d'entrer ou de rester dans un état inactif, et peut ainsi empêcher les disques durs de réduire leur vitesse de rotation lorsqu'ils ne sont pas utilisés. Les demandes excessives et non-nécessaires sur le stockage ont été minimisées dans certaines applications. En particulier, les demandes qui empêchaient les disques durs de réduire leur vitesse de rotation.

Audit initscript

Les services qui démarrent automatiquement qu'ils soient requis de le faire ou non, présentent une perte des ressources du système potentielle. Ces services devraient plutôt être par défaut « off » (éteints) ou « on demand » (à la demande) dans la mesure du possible. Par exemple, le service **BlueZ**,

qui active la prise en charge Bluetooth, se mettait auparavant automatiquement en route, dès lors que le système était allumé ; et ce, que du matériel Bluetooth soit présent ou non. L'initscript **BlueZ** vérifie maintenant que du matériel Bluetooth est présent sur le système avant de lancer ce service.



CHAPITRE 4. CAS D'UTILISATION

Ce chapitre décrit deux types de cas d'utilisation dans le but d'illustrer les méthodes d'analyse et de configuration décrites ailleurs dans ce guide. Le premier exemple prend en considération des serveurs typiques et le deuxième est un ordinateur portable commun.

4.1. EXEMPLE – SERVEUR

De nos jours, un serveur standard typique est fourni avec toutes les fonctionnalités de matériel nécessaires prises en charge dans Red Hat Enterprise Linux 6. La première chose à faire est de prendre en considération les types de charges de travail pour lesquelles le serveur va être utilisé. En vous basant sur ces informations, vous pourrez ensuite décider quels composants pourront être optimisés pour réaliser des économies d'énergie.

Peu importe le type de serveur, la performance des graphismes n'est généralement pas requise. Ainsi, les économies d'énergie GPU peuvent être laissées en cours d'utilisation.

Serveur web

Un serveur web nécessite un réseau et une E/S de disque. Selon la vitesse de connexion extérieure, 100 Mo/s peut suffire. Si la machine sert principalement à des pages statiques, la performance du CPU peut ne pas être très importante. Ainsi, les choix de gestion de l'alimentation qui s'offrent à vous incluent :

- aucun plugin de réseau ou de disque pour **tuned**.
- ALPM en marche.
- Gouverneur **ondemand** en marche.
- carte réseau limitée à 100 Mo/s.

Serveur de calcul

Un serveur de calcul nécessite principalement des CPU. Les choix de gestion de l'alimentation s'offrant à vous incluent :

- des plugins de réseau ou de disque pour **tuned** selon les tâches et l'endroit où sont stockées les données ; ou pour des systèmes en mode batch, **tuned** est totalement actif.
- selon l'utilisation, le gouverneur **performance**.

Serveur mail

Un serveur mail nécessite principalement des E/S de disque et des CPU. Les choix de gestion de l'alimentation s'offrant à vous incluent :

- Le gouverneur **ondemand** activé, car les derniers pourcentages de la performance du CPU ne sont pas importants.
- aucun plugin de réseau ou de disque pour **tuned**.
- la vitesse du réseau ne devrait pas être limitée, car les mails sont souvent internes et peuvent ainsi bénéficier d'un lien de 1 Go/s ou de 10 Go/s.

Serveur de fichiers

Les prérequis d'un serveur de fichiers sont similaires à ceux d'un serveur mail, mais selon le protocole

utilisé, celui-ci peut nécessiter une meilleure performance CPU. Typiquement, les serveurs basés sur Samba requièrent davantage de CPU que NFS, et NFS en requiert plus que iSCSI. Même ainsi, vous devriez être en mesure d'utiliser le gouverneur **ondemand**.

Serveur d'annuaire

Typiquement, un serveur d'annuaire requiert moins d'E/S de disque, particulièrement si il est équipé de suffisamment de RAM. La latence du réseau est importante, les E/S de réseau le sont moins. Prenez en considération le paramétrage de la latence du réseau à une vitesse de lien moindre, mais testez soigneusement ceci pour le réseau en question.

4.2. EXEMPLE – ORDINATEUR PORTABLE

Un autre domaine très commun dans lequel la gestion et l'économie d'énergie peut réellement faire une différence est l'ordinateur portable. Comme les ordinateurs portables, par définition, utilisent beaucoup moins d'énergie que les stations de travail ou les serveurs, la possibilité de réaliser des économies est, dans l'absolu, moindre que pour d'autres machines. Par contre, en mode batterie, toute économie peut se révéler utile pour obtenir quelques minutes de plus. Même si cette section se concentre sur les ordinateurs portables en mode batterie, vous pouvez sans aucun doute utiliser certains, ou la totalité, de ces paramètres sur une alimentation secteur.

Les économies sur des composants uniques font généralement une différence relative plus importante que celle qu'elles ont sur des stations de travail. Par exemple, une interface réseau de 1 Go/s exécutée à 100 Mo/s économisera environ 3–4 watts. Pour un serveur typique avec une consommation totale d'environ 400 watts, cette économie est d'environ 1 %. Sur un ordinateur portable consommant un total d'à peu près 40 watts, l'économie d'énergie de ce seul composant représenterait 10 % de l'énergie totale consommée.

Les optimisations spécifiques pour réaliser des économies d'énergie sur un ordinateur portable typique incluent :

- La configuration du BIOS du système afin de désactiver tout le matériel que vous n'utilisez pas. Par exemple, les ports parallèles ou sériels, lecteurs de cartes, webcams, WiFi, Bluetooth, etc.
- Atténuez l'affichage dans des environnements plus sombres, lorsque vous n'avez pas besoin d'un éclairage maximum pour lire l'écran de manière confortable. Utilisez **Système+Préférences → Gestion de l'alimentation** sur le bureau GNOME, **Lanceur d'application Kickoff+Ordinateur+Paramètres système+Avancé → Gestion de l'alimentation** sur le bureau KDE ; ou **gnome-power-manager** ou **xbacklight** dans la ligne de commande ; ou utilisez les touches de fonction sur votre ordinateur portable.
- Utilisez le profil **laptop-battery-powersave** de **tuned-adm** pour activer un ensemble de mécanismes d'économie d'énergie. Remarquez qu'il y aura un impact sur la performance et la latence du disque dur et de l'interface réseau.

De plus (ou alternativement), vous pouvez effectuer de nombreux petits ajustements aux différents paramètres du système :

- utilisez le gouverneur **ondemand** (par défaut activé dans Red Hat Enterprise Linux 6)
- activez le mode ordinateur portable (faisant partie du profil **laptop-battery-powersave**) :

```
echo 5 > /proc/sys/vm/laptop_mode
```

- augmente le délai de vidage sur disque (fait partie du profil **laptop-battery-powersave**) :

■

```
echo 1500 > /proc/sys/vm/dirty_writeback_centisecs
```

- désactive le watchdog nmi (fait partie du profil **laptop-battery-powersave**):

```
echo 0 > /proc/sys/kernel/nmi_watchdog
```

- active l'économie d'énergie AC97 (activé par défaut dans Red Hat Enterprise Linux 6):

```
echo Y > /sys/module/snd_ac97_codec/parameters/power_save
```

- active l'économie d'énergie multicoeur (fait partie du profil **laptop-battery-powersave**):

```
echo Y > /sys/module/snd_ac97_codec/parameters/power_save
```

- active l'auto-suspension USB:

```
for i in /sys/bus/usb/devices/*/power/autosuspend; do echo 1 > $i; done
```

Remarquez que l'auto-suspension USB ne fonctionne pas correctement avec tous les périphériques USB.

- active le paramètre d'alimentation minimum pour ALPM (fait partie du profil **laptop-battery-powersave**):

```
echo min_power > /sys/class/scsi_host/host*/link_power_management_policy
```

- monte un système de fichiers à l'aide de relatime (par défaut dans Red Hat Enterprise Linux 6):

```
mount -o remount,relatime point_de_montage
```

- active le meilleur mode d'économie d'énergie pour les disques durs (fait partie du profil **laptop-battery-powersave**):

```
hdparm -B 1 -S 200 /dev/sd*
```

- désactive les interrogations de CD-ROM (fait partie du profil **laptop-battery-powersave**):

```
hal-disable-polling --device /dev/scd*
```

- réduit la luminosité de l'écran à 50 ou moins, par exemple:

```
xbacklight -set 50
```

- active DPMS pour écran inactif:

```
xset +dpms; xset dpms 0 0 300
```


- réduit les niveaux de puissance Wi-Fi (fait partie du profil `laptop-battery-powersave`):

```
for i in /sys/bus/pci/devices/*/power_level ; do echo 5 > $i ; done
```

- désactive le Wi-Fi :

```
echo 1 > /sys/bus/pci/devices/*/rf_kill
```

- limite le réseau câblé à 100 Mo/s (fait partie du profil `laptop-battery-powersave`):

```
ethtool -s eth0 advertise 0x0F
```

DRAFT

ANNEXE A. ASTUCES POUR LES DÉVELOPPEURS

Tout bon manuel de programmation couvrira les problèmes liés à la mémoire et à l'allocation de mémoire et à la performance de certaines fonctions spécifiques. En développant votre logiciel, soyez conscient des problèmes qui peuvent augmenter la consommation d'énergie sur les systèmes sur lesquels le logiciel est en cours d'exécution. Même si ces considérations n'affectent pas chaque ligne de code, vous pouvez l'optimiser dans les domaines où se créent des goulots d'étranglement de performance.

Les techniques les plus souvent problématiques incluent :

- l'utilisation de threads.
- les réveils CPU inutiles et l'utilisation incorrecte des réveils. Si vous devez effectuer un réveil, faites-le en une seule fois et aussi vite que possible (en anglais, « race to idle »).
- l'utilisation de `[f]sync()` sans que cela ne soit nécessaire.
- les interrogations actives inutiles, ou l'utilisation de délais d'attente réguliers et de courtes durées. (Au lieu de cela, il vaut mieux réagir aux événements)
- une utilisation des réveils inefficace.
- un accès disque inefficace. Utilisez des mémoires-tampons de grande taille pour éviter un accès au disque trop fréquent. N'écrivez qu'un grand bloc à la fois.
- une utilisation inefficace des minuteurs. Groupez les minuteurs sur plusieurs applications (ou même sur plusieurs systèmes) si possible.
- des E/S, ou une consommation d'énergie ou utilisation de mémoire excessive (y compris les fuites mémoires)
- l'exécution de calculs inutiles.

Les sections suivantes examinent certains de ces domaines dans plus de détails.

A.1. UTILISATION DE THREADS

Il est communément admis que l'utilisation de threads permet aux applications de fonctionner mieux et plus vite, mais ce n'est pas toujours le cas.

Python

Python utilise le verrou global, Global Lock Interpreter^[1], ainsi le threading n'est utile que pour de grandes opérations d'E/S. `Unladen-swallow`^[2] est une implémentation plus rapide de Python avec laquelle vous serez probablement en mesure d'optimiser votre code.

Perl

À l'origine, les threads Perl furent créés pour des applications exécutées sur des systèmes sans bifurcation (de l'anglais, forking) (comme sur les systèmes avec système d'exploitation 32-bit 32-bit). Dans les threads Perl, les données sont copiées pour chaque thread (copie sur écriture). Par défaut, les données ne sont pas partagées, car les utilisateurs devraient pouvoir être en mesure de définir le niveau de partage des données. Pour le partage de données, le module `threads::shared` doit être inclus. Cependant, les données ne sont pas seulement copiées (copie sur écriture), mais le module crée aussi des variables attachées pour ces données, ce qui prendra encore plus de temps et sera encore plus lent.^[3]Windows

C

Les threads C partagent la même mémoire, chaque thread possède sa propre pile, et le noyau n'a pas à créer de nouveaux descripteurs de fichiers ou à allouer de nouvel espace mémoire. C peut réellement utiliser la prise en charge de plus de CPUs pour plus de threads, Ainsi, vous pouvez utiliser un langage de bas niveau tel que C ou C++ afin de maximiser la performance de vos threads. Si vous utilisez un langage de script, prenez en considération un binding avec C. Utilisez des profileurs afin d'identifier les parties de votre code performant mal. [4]

A.2. RÉVEILS

De nombreuses applications scannent les fichiers de configuration en cherchant des modifications. Dans de nombreux cas, le scan est effectué à intervalle fixe, toutes les minutes par exemple. Ceci peut poser problème, car le disque est forcé de se réveiller des arrêts de rotation. La meilleure solution est de trouver une bonne intervalle, un bon mécanisme de vérification, ou de vérifier les modifications avec `inotify` et de réagir aux événements. `Inotify` peut vérifier une variété des changements sur un fichier ou sur un répertoire.

Par exemple :

```
int fd;
fd = inotify_init();
int wd;
/* checking modification of a file - writing into */
wd = inotify_add_watch(fd, "./myConfig", IN_MODIFY);
if (wd < 0) {
    inotify_cant_be_used();
    switching_back_to_previous_checking();
}
...
fd_set rdfs;
struct timeval tv;
int retval;
FD_ZERO(&rdfs);
FD_SET(0, &rdfs);

tv.tv_sec = 5;
value = select(1, &rdfs, NULL, NULL, &tv);
if (value == -1)
    perror(select);
else {
    do_some_stuff();
}
...
```

L'avantage de cette approche est la variété des vérifications que vous pouvez effectuer.

La limitation principale est que le nombre de suivis disponibles sur un système est limité. Ce nombre peut être obtenu sur `/proc/sys/fs/inotify/max_user_watches` et même s'il peut être changé, ceci n'est pas recommandé. De plus, au cas où `inotify` échouerait, le code devra utiliser une différente méthode de vérification, ce qui devrait signifier qu'il y aura de nombreuses occurrences de `#if` `#define` dans le code source.

pour obtenir plus d'informations sur `inotify`, reportez-vous à la page `man inotify`.

A.3. FSYNC

Fsync est connu pour être une opération d'E/S coûteuse, mais ce n'est pas totalement vrai. Par exemple, reportez-vous à l'article de Theodore Ts'o *Don't fear the fsync!*^[5] et à la discussion l'accompagnant.

Firefox avait pour habitude d'appeler la bibliothèque `sqlite` à chaque fois que l'utilisateur cliquait sur un lien vers une nouvelle page. `Sqlite` appelait `fsync` et à cause des paramètres du système de fichiers (principalement `ext3` avec le mode `data-ordered`), il y avait une longue latence lorsque rien ne se passait. Ceci pouvait ainsi prendre longtemps (jusqu'à 30 secondes) si un autre processus copiait un fichier de grande taille au même moment.

Cependant, dans d'autres cas, où `fsync` n'était pas du tout utilisé, des problèmes sont survenus lors du basculement vers les systèmes de fichiers `ext4`. `Ext3` était réglé sur le mode `data-ordered`, qui vide la mémoire toutes les quelques secondes et l'enregistre sur le disque. Avec `ext4` et `laptop_mode`, l'intervalle entre les enregistrements était plus long et les données pouvaient être perdues lorsque le système était éteint de manière inattendue. Il existe maintenant un patch pour `ext4`, mais il faut tout de même prendre en considération le design de nos applications avec prudence, et utiliser `fsync` de manière appropriée.

L'exemple suivant illustrant simplement la lecture et l'écriture dans un fichier de configuration montre comment une copie de sauvegarde peut être réalisée, ou comment des données peuvent être perdues :

```
/* open and read configuration file e.g. ~/.kde/myconfig */
fd = open("./kde/myconfig", O_WRONLY|O_TRUNC|O_CREAT);
read(myconfig);
...
write(fd, bufferOfNewData, sizeof(bufferOfNewData));
close(fd);
```

Une meilleure approche aurait été comme suit :

```
open("./kde/myconfig", O_WRONLY|O_TRUNC|O_CREAT);
read(myconfig);
...
fd = open("./kde/myconfig.suffix", O_WRONLY|O_TRUNC|O_CREAT);
write(fd, bufferOfNewData, sizeof(bufferOfNewData));
fsync; /* paranoia - optional */
...
close(fd);
rename("./kde/myconfig", "./kde/myconfig~"); /* paranoia - optional */
rename("./kde/myconfig.suffix", "./kde/myconfig");
```

[1] <http://docs.python.org/c-api/init.html#thread-state-and-the-global-interpreter-lock>

[2] <http://code.google.com/p/unladen-swallow/>

[3] http://www.perlmonks.org/?node_id=288022

[4] <http://people.redhat.com/drepper/lt2009.pdf>

[5] <http://thunk.org/tytso/blog/2009/03/15/dont-fear-the-fsync/>

ANNEXE B. HISTORIQUE DES RÉVISIONS

Version 0.2-12.400 Rebuild with publican 4.0.0	2013-10-31	Rüdiger Landmann
Version 0.2-12 Rebuild for Publican 3.0	2012-07-18	Anthony Towns
Version 0.2-19 Ajout d'une section pour DeviceKit-Power – BZ#548242 Ajout d'une section pour GNOME Power Manager – BZ#548243	Tue Jul 20 2010	Rüdiger Landmann
Version 0.2-18 réécriture d'une phrase dans le chapitre "Importance"	Fri Jul 16 2010	Rüdiger Landmann
Version 0.2-17 correction d'erreurs typographiques	Thu Jul 8 2010	Rüdiger Landmann
Version 0.2-16 correction de commandes yum – BZ#611745 correction de Tuned – RT3#67837 correction d'erreurs typographiques – BZ#611428 traduction d'USV vers UPS	Thu Jul 8 2010	Rüdiger Landmann
Version 0.2-15 Incorporation des commentaires du développeur – BZ#608908 Suppression des références aux options -D et -G pour BLTK – BZ#593665	Wed Jun 30 2010	Rüdiger Landmann
Version 0.2-14 Arrangement	Tue Mar 15 2010	Rüdiger Landmann
Version 0.2-13 Description de l'outil BLTK	Mon Mar 15 2010	Rüdiger Landmann
Version 0.2-12 Description de l'outil PowerTOP	Mon Mar 15 2010	Rüdiger Landmann
Version 0.2-11 Description de l'outil Tuned	Mon Mar 15 2010	Rüdiger Landmann
Version 0.2-10 Description des audits effectués dans User Space	Mon Mar 15 2010	Rüdiger Landmann
Version 0.2-9 Restructuration de parties/chapitres/sections	Mon Mar 15 2010	Rüdiger Landmann
Version 0.2-8 Remarque de la prise en charge de C6 pour les CPUs Nehalem	Mon Mar 15 2010	Rüdiger Landmann
Version 0.2-7 Remarque d'une meilleure prise en charge de la suspension et de la reprise	Sun Mar 14 2010	Rüdiger Landmann
Version 0.2-6 Documentation du noyau Tickless	Sun Mar 14 2010	Rüdiger Landmann
Version 0.2-5	Sun Mar 14 2010	Rüdiger Landmann

Documentation d'ASPM		
Version 0.2-4 Documentation de relative	Sat Mar 13 2010	Rüdiger Landmann
Version 0.2-3 Documentation du plafonnement d'énergie	Sat Mar 13 2010	Rüdiger Landmann
Version 0.2-2 Documentation sur la gestion de l'alimentation améliorée des graphismes	Tue Mar 9 2010	Rüdiger Landmann
Version 0.2-1 Documentation de RFKill	Mon Mar 8 2010	Rüdiger Landmann
Version 0.2-0 Ajout d'une table des matières pour un guide agrandi, et marqué comme « brouillon »	Thu Dec 17 2009	Rüdiger Landmann
Version 0.1-0 updated to build on Publican 0.36 under product heading Whitepapers now	Tue Sep 23 2008	Don Domingo