

Fedora 14

Power Management Guide

Gestire il Consumo Energetico in Fedora



Fedora 14 Power Management Guide

Gestire il Consumo Energetico in Fedora

Edizione 1

Autore Don Domingo
Autore Rüdiger Landmann r.landmann@redhat.com

Copyright © 2010 Red Hat Inc. and others.

The text of and illustrations in this document are licensed by Red Hat under a Creative Commons Attribution–Share Alike 3.0 Unported license ("CC-BY-SA"). An explanation of CC-BY-SA is available at <http://creativecommons.org/licenses/by-sa/3.0/>. The original authors of this document, and Red Hat, designate the Fedora Project as the "Attribution Party" for purposes of CC-BY-SA. In accordance with CC-BY-SA, if you distribute this document or an adaptation of it, you must provide the URL for the original version.

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, MetaMatrix, Fedora, the Infinity Logo, and RHCE are trademarks of Red Hat, Inc., registered in the United States and other countries.

For guidelines on the permitted uses of the Fedora trademarks, refer to https://fedoraproject.org/wiki/Legal:Trademark_guidelines.

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.

All other trademarks are the property of their respective owners.

Questo documento spiega come gestire efficacemente, i consumi di energia in sistemi Fedora 14. Le seguenti sezioni spiegano diverse tecniche per diminuire il consumo energetico (sia nei server che nei portatili), e come ogni tecnica influenzi le prestazioni globali del sistema. *Nota: Questo documento è ancora in fase di sviluppo, soggetto a grosse modifiche e qui fornito in anteprima. Il contenuto e le istruzioni indicate non dovrebbero ritenersi complete e dovrebbero usarsi con cautela.*

Prefazione	v
1. Convenzioni del documento	v
1.1. Convenzioni tipografiche	v
1.2. Convenzioni del documento	vi
1.3. Note ed avvertimenti	vii
2. Inviateci i vostri commenti!	vii
1. Panoramica	1
1.1. Importanza della Gestione Energetica	1
1.2. Principi base di Gestione Energetica	2
2. Auditing ed analisi della Gestione Energetica	5
2.1. Panoramica su Audit e Analisi	5
2.2. PowerTOP	5
2.3. Diskdevstat e netdevstat	7
2.4. Battery Life Tool Kit	10
2.5. Tuned e ktune	12
2.5.1. Il file tuned.conf	13
2.5.2. Tuned-adm	14
2.6. UPower	16
2.7. GNOME Power Manager	17
2.8. Altri mezzi per l'auditing	17
3. Infrastruttura Centrale e Meccanismi	19
3.1. Stati inattivi della CPU	19
3.2. Usare i regolatori CPUFreq	19
3.2.1. Tipi di regolatori CPUfreq	20
3.2.2. Impostazione di CPUfreq	21
3.2.3. Parametri di regolazione di CPUfreq	22
3.3. Sospensione e Ripristino	23
3.4. Kernel Tickless	23
3.5. Active State Power Management	24
3.6. Aggressive Link Power Management	24
3.7. Ottimizzazione di accesso al disco Relatime	25
3.8. Power Capping	26
3.9. Migliore Gestione Energetica nei Sistemi Grafici	27
3.10. RFKill	27
3.11. Ottimizzazione nello Spazio Utente	28
4. Casi d'Uso	31
4.1. Caso Server	31
4.2. Caso Laptop	32
A. Suggerimenti per Sviluppatori	35
A.1. Usare i Thread	35
A.2. Wakeup	36
A.3. fsync	37
B. Cronologia Revisione	39

DRAFT

Prefazione

1. Convenzioni del documento

Questo manuale utilizza numerose convenzioni per evidenziare parole e frasi, ponendo attenzione su informazioni specifiche.

Nelle edizioni PDF e cartacea questo manuale utilizza caratteri presenti nel set *Font Liberation*¹. Il set Font Liberation viene anche utilizzato nelle edizioni HTML se il set stesso è stato installato sul vostro sistema. In caso contrario, verranno mostrati caratteri alternativi ma equivalenti. Da notare: Red Hat Enterprise Linux 5 e versioni più recenti, includono per default il set Font Liberation.

1.1. Convenzioni tipografiche

Vengono utilizzate quattro convenzioni tipografiche per richiamare l'attenzione su parole e frasi specifiche. Queste convenzioni, e le circostanze alle quali vengono applicate, sono le seguenti.

Neretto monospazio

Usato per evidenziare l'input del sistema, incluso i comandi della shell, i nomi dei file ed i percorsi. Utilizzato anche per evidenziare tasti e combinazione di tasti. Per esempio:

Per visualizzare i contenuti del file `my_next_bestselling_novel` nella vostra directory di lavoro corrente, inserire il comando `cat my_next_bestselling_novel` al prompt della shell e premere **Invio** per eseguire il comando.

Quanto sopra riportato include il nome del file, un comando della shell ed un tasto, il tutto riportato in neretto monospazio e distinguibile grazie al contesto.

Le combinazioni di tasti possono essere distinte dai tasti tramite il trattino che collega ogni parte della combinazione. Per esempio:

Premere **Invio** per eseguire il comando.

Premere **Ctrl+Alt+F2** per smistarsi sul primo virtual terminal. Premere **Ctrl+Alt+F1** per ritornare alla sessione X-Windows.

Il primo paragrafo evidenzia il tasto specifico singolo da premere. Il secondo riporta due combinazioni di tasti, (ognuno dei quali è un set di tre tasti premuti contemporaneamente).

Se si discute del codice sorgente, i nomi della classe, i metodi, le funzioni i nomi della variabile ed i valori ritornati indicati all'interno di un paragrafo, essi verranno indicati come sopra, e cioè in **neretto monospazio**. Per esempio:

Le classi relative ad un file includono **filesystem** per file system, **file** per file, e **dir** per directory. Ogni classe possiede il proprio set associato di permessi.

Proportional Bold

Ciò denota le parole e le frasi incontrate su di un sistema, incluso i nomi delle applicazioni; il testo delle caselle di dialogo; i pulsanti etichettati; le caselle e le etichette per pulsanti di selezione, titoli del menu e dei sottomenu. Per esempio:

¹ <https://fedorahosted.org/liberation-fonts/>

Selezionare **Sistema** → **Preferenze** → **Mouse** dalla barra del menu principale per lanciare **Preferenze del Mouse**. Nella scheda **Pulsanti**, fate clic sulla casella di dialogo **mouse per mancini**, e successivamente fate clic su **Chiudi** per cambiare il pulsante primario del mouse da sinistra a destra (rendendo così il mouse idoneo per un utilizzo con la mano sinistra).

Per inserire un carattere speciale in un file **gedit**, selezionare **Applicazioni** → **Accessori** → **Mappa carattere** dalla barra menu principale. Successivamente, selezionare **Cerca** → **Trova...** dalla barra del menu **Mappa carattere**, inserire il nome del carattere nel campo **Cerca** e cliccare **Successivo**. Il carattere ricercato verrà evidenziato nella **Tabella caratteri**. Fare un doppio clic sul carattere evidenziato per posizionarlo nel campo **Testo da copiare**, e successivamente fare clic sul pulsante **Copia**. Ritornare ora al documento e selezionare **Modifica** → **Incolla** dalla barra del menu di **gedit**.

Il testo sopra riportato include i nomi delle applicazioni; nomi ed oggetti del menu per l'intero sistema; nomi del menu specifici alle applicazioni; e pulsanti e testo trovati all'interno di una interfaccia GUI, tutti presentati in neretto proporzionale e distinguibili dal contesto.

Corsivo neretto monospazio o ***Corsivo neretto proporzionale***

Sia se si tratta di neretto monospazio o neretto proporzionale, l'aggiunta del carattere corsivo indica un testo variabile o sostituibile. Il carattere corsivo denota un testo che non viene inserito letteralmente, o visualizzato che varia a seconda delle circostanze. Per esempio:

Per collegarsi ad una macchina remota utilizzando ssh, digitare **ssh *username@domain.name*** al prompt della shell. Se la macchina remota è **example.com** ed il nome utente sulla macchina interessata è john, digitare **ssh *john@example.com***.

Il comando **mount -o remount *file-system*** rimonta il file system indicato. Per esempio, per rimontare il file system **/home**, il comando è **mount -o remount */home***.

Per visualizzare la versione di un pacchetto attualmente installato, utilizzare il comando **rpm -q *package***. Esso ritornerà il seguente risultato: ***package-version-release***.

Da notare la parola in Corsivo neretto — nome utente, domain.name, file-system, pacchetto, versione e release. Ogni parola racchiude il testo da voi inserito durante l'emissione di un comando o per il testo mostrato dal sistema.

Oltre all'utilizzo normale per la presentazione di un titolo, il carattere Corsivo denota il primo utilizzo di un termine nuovo ed importante. Per esempio:

Publican è un sistema di pubblicazione per *DocBook*.

1.2. Convenzioni del documento

Gli elenchi originati dal codice sorgente e l'output del terminale vengono evidenziati rispetto al testo circostante.

L'output inviato ad un terminale è impostato su **tondo monospazio** e così presentato:

```
books      Desktop  documentation  drafts  mss    photos  stuff  svn
books_tests Desktop1  downloads      images  notes  scripts  svgs
```

Gli elenchi del codice sorgente sono impostati in **tondo monospazio** ma vengono presentati ed evidenziati nel modo seguente:

```
package org.jboss.book.jca.ex1;

import javax.naming.InitialContext;

public class ExClient
{
    public static void main(String args[])
        throws Exception
    {
        InitialContext iniCtx = new InitialContext();
        Object ref = iniCtx.lookup("EchoBean");
        EchoHome home = (EchoHome) ref;
        Echo echo = home.create();

        System.out.println("Created Echo");

        System.out.println("Echo.echo('Hello') = " + echo.echo("Hello"));
    }
}
```

1.3. Note ed avvertimenti

E per finire, tre stili vengono usati per richiamare l'attenzione su informazioni che in caso contrario potrebbero essere ignorate.

Nota Bene

Una nota è un suggerimento o un approccio alternativo per il compito da svolgere. Non dovrebbe verificarsi alcuna conseguenza negativa se la nota viene ignorata, ma al tempo stesso potreste non usufruire di qualche trucco in grado di facilitarvi il compito.

Importante

Le caselle 'importante' riportano informazioni che potrebbero passare facilmente inosservate: modifiche alla configurazione applicabili solo alla sessione corrente, o servizi i quali necessitano di un riavvio prima di applicare un aggiornamento. Ignorare queste caselle non causa alcuna perdita di dati ma potrebbe causare irritazione e frustrazione da parte dell'utente.

Avvertenza

Un Avvertimento non dovrebbe essere ignorato. Se ignorato, potrebbe verificarsi una perdita di dati.

2. Inviateci i vostri commenti!

Se individuate degli errori di battitura in questo manuale, o se pensate di poter contribuire al suo miglioramento, contattateci subito! Inviare i vostri suggerimenti tramite Bugzilla: <http://bugzilla.redhat.com/bugzilla/> sul componente **Fedora Documentation**.

Quando inviate un bug report, assicuratevi di indicare l'identificatore del manuale: *power-management-guide*

Se inviate un suggerimento per contribuire al miglioramento della guida, cercate di essere il più specifici possibile. Se avete individuato un errore, indicate il numero della sezione e alcune righe di testo, in modo da agevolare la ricerca dell'errore.

DRAFT

Panoramica

Limitare l'energia usata dai sistemi di computer è uno dei principali punti del *green IT* (ambientalismo IT), un insieme di considerazioni che abbracciano anche l'uso di materiali riciclabili, l'impatto sull'ambiente della produzione hardware ed una consapevolezza ambientale a partire dalla progettazione e durante tutto il ciclo di vita dei sistemi. In questo documento vengono fornite le informazioni e le linee guida riguardo alla gestione energetica dei sistemi che eseguono Fedora 14.

1.1. Importanza della Gestione Energetica

Alla base di una gestione energetica, sta una comprensione di come ottimizzare efficacemente il consumo di ogni componente del sistema. Ciò richiede di studiare le varie attività eseguite del sistema e di configurare ogni componente assicurando prestazioni idonee al compito da svolgere.

Il motivo principale di ogni gestione energetica è:

- ridurre i consumi per tagliare i costi.

Una gestione energetica corretta produce i seguenti vantaggi:

- riduce l'emissione termica dei server e dei centri di calcolo informatici
- riduce i costi indiretti: sistemi di raffreddamento, locali, cavi, generatori, *Gruppi di continuità* o UPS (Uninterruptible Power Supplies)
- aumenta la durata media della batteria nei portatili
- riduce le emissioni di anidride carbonica
- soddisfa le norme nazionali/internazionali relative all' Green IT, per esempio Energy Star
- soddisfa le linee guida aziendali per i nuovi sistemi

In generale, una riduzione dei consumi in un componente (o nell'intero sistema), comporta un abbassamento del livello termico e quindi anche delle prestazioni. Perciò, diventa importante analizzare e testare il calo delle prestazioni, apportato dalle varie configurazioni, soprattutto in sistemi *mission-critical*.

Studiando le diverse operazioni eseguite dal sistema, e configurando ogni componente in modo da assicurare una prestazione sufficiente al compito da svolgere, sarà possibile risparmiare energia, dissipare meno calore ed ottimizzare il ciclo di vita delle batterie nei portatili. I principi base per analizzare e regolare un sistema, in funzione del consumo energetico, sono simili a quelli per regolare le prestazioni. A un certo livello, la gestione energetica e la regolazione delle prestazioni sono approcci opposti in una configurazione, poichè i sistemi generalmente sono ottimizzati o per una migliore prestazione o per una migliore gestione dei consumi. Questo manuale descrive i tool forniti dal Fedora Project e le tecniche da noi sviluppate per questo processo.

Fedora 14 viene già distribuita con varie nuove caratteristiche per la gestione dell'energia, abilitate per impostazione. Esse sono state selezionate in modo da non influire sulle prestazioni d'uso, di un server tipico o desktop. Tuttavia per usi specifici, in cui si richiedano rapidi tempi di risposta, bassa latenza o elevate prestazioni della CPU, potrebbe essere indispensabile ricontrollare le suddette impostazioni predefinite.

Per stabilire se è il caso di ottimizzare le proprie macchine usando le tecniche descritte in questo documento, si provi a rispondere alle seguenti domande:

Domanda Corre davvero ottimizzare?

Risposta Il processo di ottimizzazione dell'energia può essere una esigenza aziendale che prevede di seguire alcune linee guida o una imposizione dettata da altre norme.

Domanda Quanto occorre ottimizzare?

Risposta Se diverse tecniche quì esposte sono delle linee guida, che vogliono offrire un insieme di ottimizzazioni generali per migliorare l'impiego delle risorse energetiche, e che non richiedono di effettuare l'intero processo di audit e di analisi in tutti i dettagli, sulla propria macchina. Ovviamente, non risulteranno così precise come nei sistemi sui quali è stato eseguito un audit ed un'analisi accurata, ma offrono comunque un buon compromesso.

Domanda Il processo di ottimizzazione rischia di ridurre le prestazioni al di sotto di un livello accettabile?

Risposta La maggior parte delle tecniche descritte in questo documento, hanno un impatto diretto sulle prestazioni del sistema. Se si desidera implementare una propria gestione energetica, che vada al di là delle impostazioni predefinite già presenti in Fedora 14, si consiglia di controllare, dopo aver eseguito l'ottimizzazione, le prestazioni del sistema, per verificare se la riduzione delle prestazioni sia accettabile.

Domanda Il tempo e le risorse necessarie per ottimizzare il sistema non saranno superiori ai benefici raggiungibili?

Risposta Generalmente non vale la pena eseguire una ottimizzazione, seguendo l'intero processo, su un singolo sistema, poichè il tempo ed i costi necessari superano di gran lunga i benefici che si possono ottenere durante il ciclo di vita di una singola macchina. D'altro canto, se per esempio si utilizzano 10000 sistemi desktop i quali utilizzano la stessa impostazione e configurazione, allora l'impostazione ottimizzata e la sua implementazione su tutte e 10000 le macchine potrebbe essere una buona idea.

Le seguenti sezioni spiegano come configurazioni per prestazioni hardware ottimali, siano un beneficio per il sistema, in termini di consumi energetici.

1.2. Principi base di Gestione Energetica

Una gestione energetica efficace si basa sui seguenti principi:

Una CPU inattiva dovrebbe essere attivata solo quando necessario

Il kernel usato nelle versioni di Fedora fino a Fedora 8, usava un timer periodico per ogni CPU, che impediva la effettiva sospensione della CPU, in quanto la CPU doveva elaborare gli eventi legati al timer (che, in base alle impostazioni, potevano susseguirsi ogni pochi millisecondi), a prescindere dallo stato di esecuzione o meno dei processi. Detto ciò, una tecnica molto importante per una efficace gestione energetica, consiste nel diminuire la frequenza di attivazione della CPU.

E proprio per questo, il kernel Linux in Fedora 14 elimina il timer periodico: lo stato inattivo di una CPU è ora *tickless*, che impedisce alla CPU di consumare energia non indispensabile, durante lo stato inattivo. Tuttavia, i benefici di questo stato tickless si riducono o annullano del tutto, in presenza di applicazioni che creano eventi timer, non necessari, come per esempio i cosiddetti *polling event* (gli eventi di controllo modifica-volume o movimento-mouse e simili).

Fedora 14 include diversi tool per analizzare e identificare le applicazioni, in base al loro uso della CPU. Per i dettagli, consultare il [Capitolo 2, Auditing ed analisi della Gestione Energetica](#).

Dispositivi e hardware non utilizzati dovrebbero essere completamente disabilitati

Ciò interessa soprattutto i dispositivi con parti mobili (per esempio, gli hard disk). In aggiunta a questi, alcune applicazioni possono lasciare un dispositivo non più utilizzato, ma abilitato, in uno stato *open*; in tal caso il kernel assume che il dispositivo sia ancora in uso, impedendo così al dispositivo di entrare in uno stato di risparmio energetico.

Bassa attività dovrebbe corrispondere a basso wattaggio

In molti casi, ciò dipende da hardware moderno e da una corretta configurazione del BIOS. I componenti obsoleti, spesso non offrono supporto ad alcune delle nuove caratteristiche disponibili in Fedora 14. Assicurarsi di utilizzare firmware ufficiale aggiornato e di abilitare, nel BIOS, nelle sezioni per la gestione energetica o per la configurazione dei dispositivi, le funzioni per la gestione energetica. Alcune impostazioni da individuare sono:

- *SpeedStep*¹ : Tecnologia sviluppata da Intel per ridurre il consumo dei suoi processori
- *PowerNow!*² : Tecnologia sviluppata da AMD per ridurre il consumo nei processori dei portatili
- *Cool'n'Quiet*³ : Implementazione dinamica di AMD dell'underclocking per il risparmio di energia
- *ACPI (C state)*⁴ : Advanced Configuration and Power Interface o ACPI, fornisce tra l'altro, uno standard aperto per la gestione energetica di scheda madre e periferiche
- *S.M.A.R.T.*⁵ : Self-Monitoring, Analysis and Reporting Technology o S.M.A.R.T., è un sistema di monitoraggio per dischi rigidi, per rilevare e fornire diversi indicatori di affidabilità

Se l'hardware supporta queste funzioni ed esse sono abilitate nel BIOS, Fedora 14 le userà per impostazione.

Differenti stati di CPU e loro effetti

Le moderne CPU dotate di *ACPI* o Advanced Configuration and Power Interface, presentano diversi stati di potenza. I tre diversi stati sono:

- Sleep (C-states)
- Frequency (P-states)
- Heat output (T-states o "stati termici")

Una CPU nello stato di inattività più basso, consuma la potenza minima, ma al tempo stesso, quando richiesto, richiede un tempo maggiore per riattivarsi. In casi molto rari, ciò può portare la CPU a riattivarsi immediatamente, ogni volta sia appena entrata nello stato inattivo, con la conseguenza di una CPU in uno stato perennemente occupato e la perdita del guadagno precedentemente ottenuto in altri stati.

Una macchina spenta consuma la potenza minima

Può sembrare ovvio, ma uno dei modi più efficaci per risparmiare energia è spegnere completamente i sistemi. Per esempio, un'azienda potrebbe contribuire allo sviluppo di un *IT ecologico* consapevole, prevedendo tra le proprie linee guida, lo spegnimento delle macchine durante la pausa pranzo o dopo gli orari lavorativi. Potrebbe altresì consolidare i vari server fisici in un server più grande, magari virtualizzandoli usando le tecnologie che noi distribuiamo con Fedora 14.

DRAFT

Auditing ed analisi della Gestione Energetica

2.1. Panoramica su Audit e Analisi

Una dettagliata operazione di audit, analisi e regolazione su un sistema singolo, generalmente è più un'eccezione che una regola, poiché il tempo ed i costi spesi in tale operazione, sarebbero ben superiori rispetto ai vantaggi ottenuti da queste sofisticate messe a punto. Tuttavia, se applicate, una volta, su un numero consistente di sistemi simili, in cui sia possibile utilizzare le stesse impostazioni su tutti i sistemi, allora esse possono rivelarsi molto utili. Per esempio, si consideri uno schieramento di qualche migliaio di sistemi desktop, o un cluster HPC (High Performance Computing), dove le macchine sono quasi tutte identiche. Un altro motivo che giustifica l'auditing e l'analisi, è la possibilità di creare una base di partenza rispetto alla quale fare confronti, rilevando modifiche o regressioni nel ciclo di vita del sistema. I risultati di queste analisi possono essere molto utili nei casi in cui l'hardware, il BIOS o gli aggiornamenti software avvengano con regolarità e si desideri evitare sorprese nei consumi. Generalmente, un accurato processo di audit e di analisi dà un'idea molto migliore di cosa stia succedendo su un particolare sistema.

L'auditing e l'analisi di un sistema in funzione dei consumi energetici, è un'operazione relativamente complessa, anche nei sistemi più moderni. Numerosi sistemi non sono dotati delle necessarie interfacce per misurare via software, i propri consumi. Tuttavia non mancano le eccezioni: *Integrated Lights-Out*¹ o ILO, una console di gestione per sistemi server Hewlett Packard, ha un modulo per la gestione dei consumi accessibile via web. IBM fornisce una soluzione simile nei propri moduli per *BladeCenter*². In alcuni sistemi Dell, anche *IT Assistant*³ offre capacità di monitorare i consumi. Altri produttori offrono capacità simili per le proprie piattaforme server, ma come si può intuire non esiste una unica soluzione che sia supportata da tutti. Se il proprio sistema non ha nessun meccanismo interno per misurare i consumi, si possono usare alcune alternative possibili. Si potrebbe installare un particolare alimentatore che informi sui consumi via USB. Un esempio è Odin GT 550 W PC della Gigabyte ed il software specifico che dà lettura di questi valori in Linux è disponibile su [questa pagina](#)⁴. Altra possibilità è usare dei wattometri esterni, come Watts up? PRO forniti di connettore USB.

La misurazione diretta dei consumi diventa indispensabile quando occorre massimizzare il risparmio energetico. Comunque, sono disponibili altri mezzi per misurare gli effetti dei cambiamenti sull'andamento del sistema. Questo capitolo descrive i tool necessari.

2.2. PowerTOP

L'introduzione in Fedora del kernel *tickless*, permette alla CPU di entrare più frequentemente in uno stato inattivo, riducendo i consumi e migliorando la gestione energetica. Il nuovo tool **PowerTOP** identifica componenti specifici nelle applicazioni dello spazio utente e nel kernel, che più frequentemente attivano la CPU. Durante la fase di sviluppo, **PowerTOP** è stato usato per adempiere agli audit descritti nella [Sezione 3.11, «Ottimizzazione nello Spazio Utente»](#), e ciò ha contribuito alla regolazione di molte applicazioni, che in questa release riducono di almeno dieci volte le inutili attivazioni della CPU.

¹ http://en.wikipedia.org/wiki/HP_Integrated_Lights-Out

² http://en.wikipedia.org/wiki/IBM_BladeCenter

³ http://en.wikipedia.org/wiki/OpenManage#ITA_-_IT_Assistant

⁴ <http://mgmt.szh.hu/~andras/dev/gopsu/>

Installare **PowerTOP** con il comando:

```
yum install powertop
```

Eeguire **PowerTOP** con il comando:

```
powertop
```

Notare che occorre essere root, per poter eseguire **PowerTOP** nel pieno delle sue funzioni.

Durante la sua esecuzione, **PowerTOP** raccoglie informazioni sul sistema, creando un elenco di tutti i componenti che attivano più frequentemente la CPU. Altresì **PowerTOP** è in grado di fornire indicazioni per ridurre i consumi. Queste indicazioni sono visibili nella parte bassa della schermata e possono essere confermate premendo il tasto specificato. Poichè **PowerTOP** esegue aggiornamenti periodici, le indicazioni variano di concerto. Come si può vedere nella [Figura 2.1, «PowerTOP in funzione»](#), compare il suggerimento: **increase the VM dirty writeback time** (aumentare il tempo di dirty writeback della Memoria Virtuale), ed il tasto (**W**) per accettare il suggerimento.

Mentre è in esecuzione, **PowerTOP** raccoglie informazioni sul sistema e le rappresenta elencandole. Nella parte alta, un elenco indica l'intervallo di tempo in cui le CPU sono rimaste negli stati C e P. Più a lungo una CPU rimane nello stato C o P più alto, tanto meglio (**C4** è più alto di **C3**), in quanto è un buon indicatore della regolazione del sistema nell'utilizzo della CPU. Si dovrebbe avere la CPU nello stato C o P più alto, per il 90% o più, durante ogni periodo di inattività.

La seconda sezione è un sommario dei *wakeup* o riattivazioni al secondo della CPU. Il numero di wakeup al secondo, dà una buona misura di quanto i servizi o i dispositivi e i driver del kernel incidano sui consumi. Maggiore è il numero di wakeup al secondo maggiore è l'energia consumata.

Di seguito, **PowerTOP** fornisce una stima della potenza assorbita (se disponibile), dal sistema. Anche nei portatili, mentre sono in modalità batteria, **PowerTOP** indicherà questa stima.

Appresso alle stime di potenza, viene riportato un elenco dei componenti che più frequentemente riattivano la CPU dal suo stato di inattività. Nella parte alta si trovano quei componenti che influenzano maggiormente i consumi. Se si tratta di componenti del kernel, (il cui nome è racchiuso tra **<>**), allora i wakeup sono spesso generati da un driver specifico. La regolazione dei driver solitamente richiede modifiche al kernel, e ciò esula dagli scopi di questo documento. Invece i processi dello spazio utente che inviano segnali di wakeup sono più facili da gestire: per prima cosa, stabilire se il servizio o l'applicazione è necessaria; se non lo è, disattivarla. Per disabilitare, in modo permanente un servizio, eseguire:

```
chkconfig nome_del_servizio off
```

Per informazioni sull'attività del componente, eseguire:

```
ps -aux | grep nome_del_componente  
strace -p processid
```

Se l'output si ripete indefinitamente, allora molto probabilmente si è in presenza di un *busy-loop*. La sua risoluzione richiede di mettere mano al codice del componente, ma ciò esula dagli scopi di questa guida.

PowerTOP fornisce anche suggerimenti per ridurre i consumi nel sistema. Queste indicazioni appaiono nella parte bassa della schermata, e possono essere confermate premendo il tasto specificato. Nella [Figura 2.1, «PowerTOP in funzione»](#), notare il suggerimento: **increase the VM dirty writeback time**, ed il tasto (**W**) per accettare il suggerimento. Queste modifiche saranno attive al riavvio del sistema. **PowerTOP** mostra anche il comando esatto che viene eseguito, per

ottenere questa permanente ottimizzazione. Usando il proprio editor preferito, si può aggiungere il comando al file `/etc/rc.local`, in modo che l'ottimizzazione venga applicata ad ogni riavvio.

```

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)      simpress.bin : 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

```

Figura 2.1. PowerTOP in funzione

Per maggiori informazioni su **PowerTOP**, visitare il sito [Less Watts](http://www.lesswatts.org)⁵.

2.3. Diskdevstat e netdevstat

Diskdevstat e **netdevstat** sono strumenti di **SystemTap** che rilevano l'incidenza sulle attività del disco e della rete, di ogni applicazione in esecuzione in un sistema. Questi tool si ispirano a **PowerTOP** (vedere la [Sezione 2.2, «PowerTOP»](#)), che per ogni applicazione misura il numero di wakeup al secondo della CPU. Con queste informazioni, si possono identificare le applicazioni che con frequenti operazioni di I/O di breve durata consumano maggiore energia, rispetto a quelle con accessi meno frequenti e di maggiore durata. Altri tool che misurano soltanto la velocità di trasferimento dei dati, non sono in grado di fornire lo stesso tipo di informazioni d'uso.

Installare questi strumenti di **SystemTap**, usando il comando:

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

⁵ <http://www.lesswatts.org/projects/powertop/known.php>

Avviare i tool con il comando:

```
diskdevstat
```

o il comando:

```
netdevstat
```

Entrambi i comandi accettano fino a tre parametri:

diskdevstat *intervallo_di_aggiornamento* *durata_complessiva*
visualizza_istogramma

netdevstat *intervallo_di_aggiornamento* *durata_complessiva*
visualizza_istogramma

intervallo_di_aggiornamento

Il periodo in secondi per l'aggiornamenti dei dati. Per impostazione: **5**

durata_complessiva

Il periodo di esecuzione in secondi. Per impostazione predefinita: **86400** (1 giorno)

visualizza_istogramma

Flag per specificare di visualizzare i dati raccolti in istogrammi (usare un qualsiasi carattere).

L'output è molto simile a quello di **PowerTOP**. Di seguito si riporta un esempio di una lunga esecuzione di **diskdevstat** su un sistema Fedora 10, con l'ambiente KDE 4.2:

PID	UID	DEV	WRITE_CNT	WRITE_MIN	WRITE_MAX	WRITE_AVG	READ_CNT	READ_MIN	READ_MAX
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							

Le colonne indicano i seguenti dati:

PID

l'ID di processo dell'applicazione

UID

l'ID utente proprietario delle applicazioni in esecuzione

DEV

il dispositivo sul quale si è avuto l'I/O

WRITE_CNT

il numero totale di operazioni di scrittura

WRITE_MIN

il tempo minimo tra due operazioni consecutive di scrittura (in secondi)

WRITE_MAX

il tempo massimo tra due operazioni consecutive di scrittura (in secondi)

WRITE_AVG

il tempo medio tra due operazioni consecutive di scrittura (in secondi)

READ_CNT

il numero totale di operazioni di lettura

READ_MIN

il tempo minimo tra due operazioni consecutive di lettura (in secondi)

READ_MAX

il tempo massimo tra due operazioni consecutive di lettura (in secondi)

READ_AVG

il tempo medio tra due operazioni consecutive di lettura (in secondi)

COMMAND

il nome del processo

In questo esempio, si risaltano in particolar modo, tre applicazioni:

PID	UID	DEV	WRITE_CNT	WRITE_MIN	WRITE_MAX	WRITE_AVG	READ_CNT	READ_MIN	READ_MAX
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							

Esse presentano un **WRITE_CNT** maggiore di **0**; ciò significa che durante la misurazione si è verificato una qualche forma di scrittura, con **plasma** che fa registrare il maggior numero di operazioni di scrittura, e di conseguenza il minor tempo medio tra due operazioni di scrittura. Quindi, in un'ottica concernente le inefficienze energetiche delle applicazioni, **plasma** sarebbe il candidato ideale su cui investigare.

Usando i comandi **strace** ed **ltrace**, che analizzano tutte le chiamate di sistema di un dato processo (nell'esempio PID=2789), si è potuto esaminare **plasma** in maniera più approfondita. Nel nostro caso:

```
strace -p 2789
```

In tal caso, l'output di **strace** presentava ogni 45 secondi, un insieme di righe ripetitive, in cui veniva aperto in scrittura ed immediatamente richiuso, il file dell'icona KDE dell'utente. Tale attività, comportava una indispensabile operazione di scrittura sul disco, poichè i metadati del file (in particolare il timestamp), venivano modificati. La correzione finale, è stata quella di impedire le chiamate non necessarie, nel nostro caso le chiamate in assenza di aggiornamenti reali nelle icone.

2.4. Battery Life Tool Kit

Battery Life Tool Kit (BLTK) è un kit di strumenti usato per simulare ed analizzare le prestazioni ed il ciclo di vita di una batteria. BLTK in effetti, simula le attività di gruppi di utenti specifici. Sviluppato per testare le prestazioni di notebook, BLTK è in grado di analizzare anche le prestazioni di computer desktop quando avviato con l'opzione **-a**.

BLTK permette di generare carichi di lavoro riproducibili che sono comparabili con casi d'uso reali. Per esempio, il carico di lavoro **office** scrive un testo, effettua modifiche al testo, e fa lo stesso con uno spreadsheet. Usando BLTK in combinazione con **PowerTOP** o altro tool di auditing o di analisi, si potrà verificare se le ottimizzazioni abbiano avuto l'effetto desiderato quando la macchina è in piena attività. Poichè è possibile eseguire lo stesso carico di lavoro con impostazioni differenti, ripetendo il test più volte, sarà possibile confrontare i risultati delle diverse impostazioni.

Installare BLTK con il comando:

```
yum install bltk
```

Eeguire BLTK con il comando:

```
bltk carico_di_lavoro opzioni
```

Per esempio, per eseguire il carico di lavoro **idle** per 120 secondi:

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

I carichi di lavoro disponibili per impostazione, sono:

-I, --idle

il sistema è sospeso, da usare come base per confronto con altri carichi di lavoro

-R, --reader

simula la lettura di documenti (per impostazione, usa **Firefox**)

-P, --player

simula la visione di file multimediali dal drive CD/DVD (per impostazione, usa **mplayer**)

-O, --office

simula la modifica di documenti di **OpenOffice.org**

Altre opzioni possibili sono:

-a, --ac-ignore

Ignora la disponibilità dell'alimentazione di rete (AC) (indispensabile nei sistemi desktop, un pò meno nei portatili)

-T numero_di_secondi, --time numero_di_secondi

il periodo (in secondi) entro cui eseguire il test; usare questa opzione con il carico di lavoro **idle**

-F nome_del_file, --file nome_del_file

specifica il file da usare con un carico di lavoro; per esempio, con il carico di lavoro **player**, riproduce il file invece di accedere al lettore CD/DVD

-W applicazione, --prog applicazione

specifica l'applicazione da usare con un carico di lavoro; per esempio, con il carico di lavoro **reader**, specifica un browser diverso da **Firefox**

BLTK supporta un gran numero di opzioni molto specializzate. Per informazioni, consultare le pagine di man su **bltk**.

BLTK salva i risultati generati, nella directory specificata nel file di configurazione **/etc/bltk.conf** — per impostazione predefinita, in **~/ .bltk/workload.results.number/**. Per esempio, la directory **~/ .bltk/reader.results.002/** contiene i risultati del terzo test, effettuati con il carico di lavoro **reader** (il primo test è privo di numero). I risultati si trovano in vari file di testo. Per condensare i risultati in un formato di facile lettura, eseguire:

```
bltk_report path_alla_directory_con_i_risultati
```

Ora i risultati, si trovano riassunti nel file **Report**, nella directory contenente i risultati. Per visualizzare i risultati in un emulatore di terminale, usare l'opzione **-o**:

```
bltk_report -o path_della_directory_con_i_risultati
```

2.5. Tuned e ktune

Tuned è un demone che analizza l'utilizzo dei componenti nel sistema ed in base alle informazioni raccolte, regola dinamicamente le impostazioni del sistema. La regolazione dinamica avviene tenendo conto del diverso modo di utilizzo dei vari componenti durante il funzionamento del sistema. Per esempio, il disco rigido è usato in modo preponderante durante il processo di boot e di creazione della sessione, ma molto meno successivamente, quando per esempio l'utente usa applicazioni come client email o OpenOffice. Analogamente, i dispositivi di rete e la CPU sono usati in modo diverso in periodi diversi. **Tuned** analizza l'attività di questi componenti, riadattando la loro attività in base al loro grado di utilizzo.

Come esempio pratico, si consideri una workstation da lavoro. La maggior parte del tempo, l'interfaccia di rete Ethernet sarà inattiva: pochissime email verranno inviate e ricevute o alcune pagine web verranno caricate. Con questi carichi di lavoro, la scheda di rete non ha sempre bisogno della massima velocità, come impostato. **Tuned** ha un plugin di monitoraggio e di regolazione per schede di rete, in grado di rilevare basse attività di rete ed automaticamente ridurre la velocità dell'interfaccia, riducendo di conseguenza i consumi. Se ad un certo istante, l'attività dell'interfaccia aumenta e perdura per un certo periodo di tempo, per esempio perchè si scarica l'immagine di un DVD o si apre una email con un allegato di molti MB, allora **Tuned** rileva tale variazione ed imposta la velocità dell'interfaccia al valore massimo, in modo da assicurare la migliore prestazione possibile. E lo stesso principio è usato anche nei plugin di monitoraggio/regolazione della CPU e dei dischi.

Per impostazione, le schede di rete non sono configurate per comportarsi come indicato, in quanto i cambiamenti di velocità possono richiedere diversi secondi, influenzando direttamente l'esperienza utente. Simili considerazioni si applicano anche ai plugin della CPU e del disco. Quando un disco viene rallentato, possono passare diversi secondi prima che il disco raggiunga di nuovo la velocità di regime operativo, e durante questo periodo si noterebbero dei fastidiosi ritardi nel sistema. Nel caso della CPU, l'effetto latenza è più contenuto ma pur sempre misurabile, anche se quasi inavvertito dall'utente.

Oltre a **tuned**, noi offriamo anche **ktune**. **ktune**, in origine era un servizio ed un framework per ottimizzare le prestazioni di una macchina dedicata. Da allora, **ktune** è stato migliorato a tal punto da far parte integrante del nostro framework di regolazione generale. Esso viene usato soprattutto nei vari profili predefiniti, descritti nella [Sezione 2.5.2, «Tuned-adm»](#).

Per installare il pacchetto *tuned* e gli script di **systemtap** associati, usare il comando:

```
yum install tuned
```

L'installazione del pacchetto *tuned*, imposta anche un semplice file di configurazione `/etc/tuned.conf`, attivando il profilo predefinito.

Avviare **tuned** con il comando:

```
service tuned start
```

Per avviare **tuned** ad ogni boot, usare:

```
chkconfig tuned on
```

tuned, presenta diverse opzioni:

-d, --daemon

avvia `tuned` come demone (in background)

-c, --conffile

usa un file di configurazione, specificando nome e percorso; per esempio `--conffile=/etc/tuned2.conf`. L'impostazione predefinita è `/etc/tuned.conf`.

-D, --debug

visualizza i messaggi di log

2.5.1. Il file `tuned.conf`

Il file `tuned.conf` contiene le impostazioni di configurazione di `tuned`. Per default, si trova in `/etc/tuned.conf`, ma avviando `tuned` con l'opzione `--conffile`, è possibile specificare un nome ed una locazione diversa.

Il file di configurazione deve contenere una sezione `[main]`, in cui sono definiti i parametri generali di `tuned`. Poi, il file contiene una sezione per ogni plugin.

La sezione `[main]` contiene le seguenti opzioni:

interval

specifica l'intervallo, in secondi, in cui `tuned` analizza e regola il sistema. Il valore predefinito è **10**.

verbose

specifica se l'output deve essere più informativo. Il valore predefinito è **False**.

logging

specifica la priorità minima dei messaggi di log. In ordine decrescente, i valori possibili sono: **critical**, **error**, **warning**, **info** e **debug**. Il valore predefinito è **info**.

logging_disable

specifica la priorità massima dei messaggi di log; i messaggio con questo livello di priorità o con priorità inferiore non vengono registrati. In ordine decrescente i valori possibili sono: **critical**, **error**, **warning**, **info** e **debug**. Il valore **notset** disabilita questa opzione.

Ogni plugin ha la propria sezione, con il nome del plugin racchiuso tra parentesi quadre; per esempio: `[CPUTuning]`. Ogni plugin può avere le proprie opzioni, ma le seguenti si applicano a tutti i plugin:

enabled

specifica se un plugin è abilitato. Il valore predefinito è **True**.

verbose

specifica se l'output deve essere più informativo. Se non impostato, il valore viene ereditato dalla sezione `[main]`.

logging

specifica la priorità minima dei messaggi di log. Se non impostato, il valore viene ereditato da `[main]`.

Di seguito, si riporta un esempio di file `tuned.conf`:

```
[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

Spesso, l'audit e l'analisi dettagliata di un sistema, può richiedere un tale dispendio di risorse umane e materiali/temporali da superare i benefici dei watt risparmiati. Nel passato, l'unica alternativa era utilizzare le impostazioni predefinite. Come alternativa a queste due situazioni estreme, Fedora 14 include profili separati per specifici casi d'uso, insieme con lo strumento **tuned-adm** per consentire di passare facilmente da un profilo all'altro, usando un terminale. Fedora 14 include un numero di profili predefiniti per tipici casi d'uso, che è possibile selezionare ed attivare usando il comando **tuned-adm**, e con cui è possibile anche crearne di nuovi, modificarli o cancellarli.

Per elencare i profili disponibili ed identificare l'attuale profilo attivo, eseguire:

```
tuned-adm list
```

Per visualizzare solo il profilo correntemente attivo, eseguire:

```
tuned-adm active
```

Per passare ad un altro profilo disponibile, eseguire:

```
tuned-adm profile nome_profilo
```

per esempio:

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

Per disabilitare tutte le regolazioni:

```
tuned-adm off
```

Alla prima installazione di **tuned**, sarà attivo il profilo **default**. Fedora 14 include anche i seguenti profili predefiniti:

default

il profilo predefinito per il risparmio energetico. Tra i profili predefiniti, è quello con la più bassa influenza sul risparmio energetico, abilitando solo i plugin di regolazione della CPU e del disco.

desktop-powersave

un profilo rivolto ai sistemi desktop. Abilita il sistema ALPM degli host SATA (vedere la [Sezione 3.6, «Aggressive Link Power Management»](#)), per la riduzione dei consumi, e i plugin di regolazione di **tuned** del disco, della scheda ethernet e della CPU.

server-powersave

un profilo rivolto ai sistemi server. Abilita il sistema ALPM degli host SATA, disabilita il *CD-ROM polling* (il controllo sulla presenza di CD/DVD, vedere le pagine di man su `hal-disable-polling`) via **HAL** ed attiva i plugin di regolazione di **tuned** del disco e della CPU.

laptop-ac-powersave

un profilo a media influenza sulla gestione energetica, rivolto ai portatili connessi alla rete elettrica. Abilita il sistema ALPM degli host SATA ed il risparmio energetico nei sistemi WiFi, oltre ai plugin di regolazione di **tuned** del disco, della scheda Ethernet e della CPU.

laptop-battery-powersave

un profilo ad alta influenza sulla gestione energetica, rivolto ai portatili alimentati a batteria. Esso attiva tutti i meccanismi dei precedenti profili, abilita il risparmio energetico nei sistemi multi-core per bassi wakeup di sistema, e si assicura che sia attivo il regolatore *ondemand* ed abilitato il risparmio energetico nel sistema audio AC97. Questo profilo può essere usato per massimizzare il risparmio di energia non solo nei portatili alimentati a batteria, ma su qualsiasi sistema. Il rovescio della medaglia di questo profilo è rappresentato da un notevole impatto sulle prestazioni, in particolare sulla latenza di I/O del disco e di rete.

throughput-performance

un profilo server per l'ottimizzazione delle prestazioni per carico tipico. Esso disabilita i meccanismi di risparmio propri di **tuned** e **ktune** ed abilita le impostazioni di **sysctl**, migliorando le prestazioni di I/O di rete e del disco e passando il controllo dell'I/O al sistema **deadline scheduler**.

latency-performance

un profilo server per l'ottimizzazione delle prestazioni per latenza tipica. Esso disabilita i meccanismi di risparmio propri di **tuned** e **ktune** ed abilita le impostazioni di **sysctl**, riducendo le latenze di I/O di rete e del disco.

Tutti i profili si trovano all'interno di sottodirectory separate, in **/etc/tune-profiles**. In tal modo, la directory **/etc/tune-profiles/desktop-powersave** contiene tutti i file e le impostazioni del profilo **desktop-powersave**. Ogni directory contiene un massimo di quattro file:

tuned.conf

la configurazione da attivare per questo profilo.

sysctl.ktune

le impostazioni di **sysctl** usate da **ktune**. Il formato delle impostazioni è identico al file **/etc/sysconfig/sysctl** (vedere le pagine man su **sysctl** e **sysctl.conf**).

ktune.sysconfig

il file di configurazione di **ktune**, che generalmente si trova in **/etc/sysconfig/ktune**.

ktune.sh

uno script di shell, in stile **init**, usato da **ktune** durante l'avvio del sistema, per la regolazione del sistema.

Il modo più semplice per creare un profilo è partire da uno già esistente. Il profilo **laptop-battery-powersave** contiene un gran numero di regolazioni, e quindi può essere una base da cui iniziare. Copiare l'intera directory nel nuovo profilo (myprofile), come indicato di seguito:

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

Modificare, secondo le proprie esigenze, i file nel nuovo profilo. Per esempio, se si desidera rilevare regolarmente la presenza di CD/DVD nel drive, commentare la riga relativa all'ottimizzazione corrispondente, nello script di **ktune.sh**, come indicato:

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

2.6. UPower

A partire da Fedora 11, **DeviceKit-power** ha inglobato sia le funzioni di gestione energetica di **HAL** sia quelle che rientravano in **GNOME Power Manager**, in precedenti versioni di Fedora (vedere la [Sezione 2.7, «GNOME Power Manager»](#)). Con Fedora 13, **DeviceKit-power** è stato ribattezzato **UPower**, che dispone di un demone, una API e di una serie di comandi da terminale. Ogni sorgente di alimentazione presente nel sistema, è rappresentato come un dispositivo; p.e. sia una batteria di un portatile, sia la rete elettrica sono entrambi rappresentati come dei dispositivi.

Il comando **upower** accetta le seguenti opzioni:

--enumerate, -e

visualizza un *object path* per ciascuna dispositivo d'alimentazione nel sistema, per esempio:

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

--dump, -d

visualizza i parametri di tutti i dispositivi di alimentazione presenti nel sistema.

--wakeups, -w

visualizza il numero di *wakeup* al minuto della CPU

--monitor, -m

analizza il sistema, controllando i cambiamenti nei dispositivi di alimentazione, per esempio il collegamento/scollegamento dalla rete elettrica, o lo stato di esaurimento della batteria. Per terminare il monitoraggio, premere **Ctrl+C**.

--monitor-detail

Rispetto all'opzione precedente, visualizza maggiori dettagli.

--show-info object_path, -i object_path

visualizza tutte le informazioni disponibili su un *object path*. Per esempio, per informazioni su una batteria rappresentata dall'*object path* `/org/freedesktop/UPower/devices/battery_BAT0`, eseguire:

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

2.7. GNOME Power Manager

GNOME Power Manager è un demone dell'ambiente desktop GNOME. A partire da Fedora 11, molte funzioni di gestione energetica, incluse in **GNOME Power Manager** in precedenti versioni di Fedora, sono state integrate in **DeviceKit-power**, rinominato **UPower** in Fedora 13 (vedere la [Sezione 2.6, «UPower»](#)). Comunque, **GNOME Power Manager** rimane il front-end per questa funzionalità. Attraverso un applet del *system-tray*, **GNOME Power Manager** notifica i cambiamenti di stato di alimentazione, nel sistema; per esempio, lo stato di esaurimento della batteria o il collegamento/ scollegamento del cavo dalla rete elettrica.

GNOME Power Manager permette anche di configurare alcune impostazioni, di base, di gestione energetica. Per accedere alle impostazioni, *right-click* su **GNOME Power Manager** nel *system-tray*, e poi selezionare **Preferenze**.

La finestra delle **Preferenze di Gestione Energetica**, contiene tre schede:

- **Alimentazione AC**
- **Alimentazione a batteria**
- **Generale**

Usare le schede **Alimentazione AC** e **Alimentazione a batteria**, per indicare dopo quanto tempo di inattività, spegnere il display e/o mettere il sistema in uno stato di *sleep* e/o rallentare la velocità di rotazione dei dischi. La scheda **Alimentazione a batteria** serve anche ad impostare la luminosità del display ed a selezionare una azione da fare, quando la batteria è quasi scarica. Per esempio, per impostazione, **GNOME Power Manager** iberna il sistema, quando la batteria raggiunge livelli molto bassi. Usare la scheda **Generale** per impostare i pulsanti di sospensione e di alimentazione sul proprio sistema, e per specificare in quali circostanze deve apparire l'icona **GNOME Power Manager**, nel *system-tray*.

2.8. Altri mezzi per l'auditing

Fedora 14 offre vari tool con cui effettuare auditing e analisi di sistema. La maggior parte di essi possono essere usati per ottenere informazioni supplementari, nel caso si voglia verificare un risultato già noto o nel caso occorrono informazioni più dettagliate su alcune parti del sistema. Molti di questi tool sono usati anche per regolare le prestazioni. Tra di essi si annoverano:

vmstat

vmstat restituisce informazioni dettagliate su processi, memoria, paging, interruzioni di I/O, trap e sull'attività della CPU. Torna utile per avere informazioni più approfondite sulle attività del sistema e sui punti in cui è occupato.

ioostat

ioostat è simile a **vmstat**, ma limitato alle operazioni di I/O dei dispositivi a blocchi, di cui fornisce maggiori informazioni.

blktrace

blktrace è un programma che traccia in modo dettagliato l'I/O dei dispositivi a blocchi, suddividendo le informazioni in singoli blocchi, ciascuno associato alla propria applicazione. È molto utile in combinazione con **diskdevstat**.

DRAFT

Infrastruttura Centrale e Meccanismi

3.1. Stati inattivi della CPU

Le CPU con architettura x86, supportano vari stati in cui alcune parti della CPU vengono disattivate o in cui le operazioni vengono eseguite con prestazioni più basse. Questi stati, noti come *stati C*, permettono ai sistemi di risparmiare energia, disattivando parzialmente la CPU non utilizzata. Gli stati, sono enumerati da 0 in ordine crescente, con i numeri più alti ad indicare minori funzionalità attive nella CPU ed un risparmio energetico maggiore. In generale, gli stati con un dato numero, hanno delle caratteristiche più o meno simili tra famiglie di processori, fermo restando i dettagli che possono variare. I quattro stati, C0-C3, sono definiti come segue:

C0

La CPU è in stato operativo o di esecuzione, in piena attività.

C1, Halt

La CPU è in uno stato in cui non esegue nessuna istruzione, ma non si trova ancora in uno stato di minima alimentazione. Nello stato C1, la CPU può riprendere l'elaborazione praticamente *senza latenza*. Tutti i processori con stati C, devono supportare questo stato. I processori Pentium 4 supportano uno stato C1, chiamato C1E, ossia uno stato per bassi consumi.

C2, Stop-Clock

La CPU è in uno stato in cui il clock è sospeso; lo stato dei registri e delle cache viene congelato, per consentire, al riavvio del clock, la ripresa dell'elaborazione, dal contesto in cui si è avuta l'interruzione. Questo è uno stato opzionale.

C3, Sleep

La CPU è in uno stato di inattività, in cui non occorre mantenere aggiornata la cache. La riattivazione da questo stato, richiede un tempo considerevolmente più lungo rispetto alla riattivazione dallo stato C2. Anche questo è uno stato opzionale.

Recenti CPU Intel con microarchitettura "Nehalem" hanno un ulteriore stato, C6, in grado di ridurre la tensione di alimentazione fin quasi a zero, con una riduzione dei consumi energetici che varia tra l'80 ed il 90%. Il kernel in Fedora 14 include ottimizzazioni per questo nuovo stato.

3.2. Usare i regolatori CPUfreq

Uno dei modi più efficaci per ridurre il consumo energetico e gli output termici del sistema, è di intervenire sulla frequenza di CPU, usando CPUfreq. CPUfreq — detto anche regolatore di velocità della CPU — permette di calibrare la velocità di clock del processore, al volo. Ciò consente al sistema di girare a velocità di clock più basse, in modo da ridurre i consumi energetici. Il regolatore CPUfreq determina sia l'entità dell'intervento (aumentare/diminuire la frequenza), sia l'istante in cui conviene cambiare frequenza.

Il regolatore controlla l'alimentazione elettrica della CPU, che influenza direttamente le prestazioni della CPU. Ogni regolatore ha un suo comportamento, scopo e idoneità in termini di carico di lavoro. Questa sezione descrive come selezionare e configurare un regolatore di frequenza o CPUfreq, esponendo le sue caratteristiche ed il tipo di carico adeguato.

Le problematiche principali relative alla gestione energetica sono:

- Ridurre il surriscaldamento dei server
- Aumentare il ciclo di vita della batteria dei portatili

In generale, una riduzione dei consumi in un componente (o nell'intero sistema), comporta un abbassamento del livello termico e quindi anche delle prestazioni. Perciò, diventa importante analizzare e testare il calo delle prestazioni, apportato dalle varie configurazioni, soprattutto in sistemi *mission-critical*.

Le seguenti sezioni spiegano come prestazioni hardware ottimali siano in grado di beneficiare un sistema, in termini di riduzione dei consumi energetici.

3.2.1. Tipi di regolatori CPUfreq

Questa sezione elenca e descrive i diversi tipi di regolatori di frequenza (CPUfreq) disponibili in Fedora 14.

cpufreq_performance

Il regolatore di Prestazioni forza la CPU ad usare la frequenza di clock più elevata possibile. Questa frequenza viene impostata staticamente, ossia non verrà più modificata. In tal modo, questo tipo di regolatore non offre *alcun beneficio al risparmio energetico*. Tale impostazione è adatta solo in presenza di prolungati carichi di lavoro molto elevati, con la CPU raramente (o quasi mai) in uno stato inattivo.

cpufreq_powersave

Al contrario del regolatore precedente, il regolatore Powersave forza la CPU ad usare la frequenza più bassa possibile. Questa frequenza viene impostata staticamente, ossia non verrà più modificata. In tal modo questo regolatore offre il massimo risparmio energetico, a costo però di una CPU dalle *minime prestazioni*.

Il termine "powersave" può ingannare, perchè una CPU a pieno carico ed a bassa velocità può consumare di più di una CPU senza carico e più veloce. Quindi, se è opportuno impostare la CPU ad usare il regolatore Powersave durante i supposti periodi di bassa attività, l'arrivo di un inaspettato aumento di carico può causare in realtà, a consumare un quantità di energia maggiore.

In altre parole, il regolatore Powersave è più un *limitatore di velocità* che non un *power saver*. Il suo utilizzo, è adatto in sistemi ed ambienti dove il surriscaldamento potrebbe essere un problema.

cpufreq_ondemand

Il regolatore Ondemand è un regolatore dinamico che imposta la frequenza della CPU al valore massimo quando il carico di lavoro è alto, ed al valore minimo quando il sistema è inattivo. Ciò permette al sistema di regolare il proprio consumo energetico in base al carico, a spese, tuttavia, di *latenze* tra i cambi di carico. Infatti, se il sistema oscilla troppo spesso tra stati con carichi di lavoro elevati e stati di inattività, la latenza potrebbe annullare ogni beneficio di risparmio-energetico/prestazioni, offerto dal regolatore Ondemand.

Per la maggior parte dei sistemi, il regolatore Ondemand è in grado di fornire il miglior compromesso tra emissione termica, consumo energetico, prestazioni e maneggevolezza. Se il sistema è occupato solo in periodi specifici della giornata, il regolatore Ondemand regolerà automaticamente la frequenza al valore massimo o minimo, in base al carico, senza alcun intervento ulteriore.

cpufreq_userspace

Il regolatore dello Spazio Utente, permette ai programmi in esecuzione nello spazio utente (o a processi in esecuzione come root) di impostare la frequenza della CPU. Questo regolatore viene normalmente usato insieme con il demone **cpuspeed**. Tra tutti i regolatori, quello dello Spazio Utente

è il più personalizzabile ed, in base alla sua configurazione, è in grado di fornire il miglior rapporto prestazione/consumo per il sistema.

cpufreq_conservative

Come il regolatore Ondemand, il regolatore Conservativo adatta la frequenza clock in base all'uso. Tuttavia, mentre il regolatore Ondemand è più drastico, cioè passa direttamente dal valore massimo al valore minimo e viceversa, il regolatore Conservativo seleziona le frequenze in maniera più graduale.

Ciò significa che il regolatore Conservativo adatta la frequenza di clock al carico e non semplicemente, su una frequenza massima o minima. Tale comportamento fornisce un risparmio significativo di energia, ma al tempo stesso è soggetto ad una *maggiore latenza* rispetto al regolatore Ondemand.

Nota

È possibile abilitare un regolatore usando il processo **cron**. Ciò consente di impostare, automaticamente, regolatori specifici durante determinati periodi del giorno. Infatti, si potrebbe per esempio, impostare un regolatore a bassa-frequenza durante i periodi di inattività (per esempio dopo l'orario lavorativo), e un regolatore ad alta-frequenza durante i periodi di carico intenso.

Vedere la [Procedura 3.2, «Abilitare un regolatore di CPUfreq»](#) per maggiori informazioni, nella [Sezione 3.2.2, «Impostazione di CPUfreq»](#).

3.2.2. Impostazione di CPUfreq

Prima di selezionare e configurare un regolatore CPUfreq è necessario aggiungere i driver di CPUfreq, appropriati.

Procedura 3.1. Come aggiungere un driver di CPUfreq

1. Usare il seguente comando per visualizzare i driver di CPUfreq disponibili nel sistema:

```
ls /lib/modules/[versione_del_kernel]/kernel/arch/[architettura]/kernel/cpu/cpufreq/
```

2. Usare **modprobe** per aggiungere il driver appropriato.

```
modprobe [driver_di_CPUfreq]
```

In questo comando, assicurarsi di non digitare anche il suffisso **.ko**, nel filename



Importante

Per la scelta di un driver di CPUfreq, si consiglia di selezionare sempre **acpi-cpufreq** invece di **p4-clockmod**. Infatti l'utilizzo del driver **p4-clockmod** riduce la frequenza di clock della CPU, ma non riduce il voltaggio. Il driver **acpi-cpufreq**, al contrario, riduce anche il voltaggio insieme alla frequenza di clock della CPU, riducendo il consumo energetico e termico, per ogni riduzione unitaria di performance.

3. Una volta impostato il driver di CPUfreq, è possibile visualizzare il regolatore correntemente usato, eseguendo il comando:

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

Per visualizzare i regolatori disponibili per una specifica CPU, digitare:

```
cat /sys/devices/system/cpu/[ID_di_cpu]/cpufreq/scaling_available_governors
```

Alcuni regolatori di CPUfreq potrebbero non essere disponibili per l'uso. In questo caso, usare **modprobe** per aggiungere i moduli necessari del kernel che abilitano il regolatore di CPUfreq da usare. Questi moduli del kernel si trovano in `/lib/modules/[versione_del_kernel]/kernel/drivers/cpufreq/`.

Procedura 3.2. Abilitare un regolatore di CPUfreq

1. Se un regolatore specifico non è al momento disponibile, usare il comando **modprobe** per abilitare il regolatore da usare. Per esempio, se si tratta del regolatore **ondemand**, digitare il seguente comando:

```
modprobe cpufreq_ondemand
```

2. Una volta caricato, è possibile abilitarlo usando:

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

3.2.3. Parametri di regolazione di CPUfreq

Una volta scelto un regolatore di CPUfreq, è possibile regolare la velocità di ogni CPU usando i parametri di regolazione nel file `/sys/devices/system/cpu/[ID_di_cpu]/cpufreq/`. I parametri sono:

- **cpuinfo_min_freq** — Mostra la frequenza operativa minima della CPU (in KHz).
- **cpuinfo_max_freq** — Mostra la frequenza operativa massima della CPU (in KHz).
- **scaling_driver** — Mostra il driver di CPUfreq usato per regolare la frequenza della CPU.
- **scaling_available_governors** — Visualizza i regolatori di CPUfreq disponibili nel kernel. Se occorre un regolatore di CPUfreq non presente in elenco, vedere la [Procedura 3.2, «Abilitare un regolatore di CPUfreq»](#) nella [Sezione 3.2.2, «Impostazione di CPUfreq»](#).
- **scaling_governor** — Mostra il regolatore di CPUfreq attualmente in uso. Per usare un regolatore diverso, eseguire `echo [regolatore] > /sys/devices/system/cpu/[ID_di_cpu]/cpufreq/scaling_governor` (vedere la [Procedura 3.2, «Abilitare un regolatore di CPUfreq»](#) nella [Sezione 3.2.2, «Impostazione di CPUfreq»](#), per maggiori informazioni).
- **cpuinfo_cur_freq** — Mostra la frequenza corrente della CPU (in KHz).
- **scaling_available_frequencies** — Elenca le frequenze possibili per la CPU, in KHz.
- **scaling_min_freq** e **scaling_max_freq** — Impostano le due frequenze *limite* della CPU, in KHz.



Importante

Quando si impostano le frequenze limite, impostare prima il valore di **scaling_max_freq** e poi quello di **scaling_min_freq**.

- **affected_cpus** — Elenca le CPU che necessitano di un software di coordinamento della frequenza.
- **scaling_setspeed** — Usato per modificare la frequenza di clock, in KHz. Si può impostare solo una frequenza compresa nei limiti fissati da **scaling_min_freq** e **scaling_max_freq**.

Per visualizzare il valore corrente di un parametro, usare **cat [parametro]**. Per esempio, per sapere la frequenza corrente (in KHz) della cpu0, usare:

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

Per modificare il valore di un parametro, usare **echo [valore] > /sys/devices/system/cpu/[ID_di_cpu]/cpufreq/[parametro]**. Per esempio, per impostare la velocità minima della cpu0 su 360 KHz, usare:

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

3.3. Sospensione e Ripristino

Quando un sistema viene temporaneamente sospeso, il kernel salva lo stato dei driver e poi li scarica. Quando il sistema viene riattivato, il kernel ricarica i driver con le impostazioni di stato salvate, cercando di ripristinare lo stato precedente alla sospensione. La capacità dei driver di soddisfare a questo compito determina il corretto ripristino del sistema.

Dal punto di vista delle funzioni di sospensione/ripristino, i driver video sono particolarmente problematici, in quanto le specifiche *ACPI* (Advanced Configuration and Power Interface), non richiedono che il firmware del sistema debba essere capace di riprogrammare l'hardware video. Per questo, a meno che i driver video non siano in grado di programmare l'hardware da uno stato completamente non-inizializzato, in generale essi non assicurano funzioni di ripristino da stati di sospensione.

Fedora 14 include miglior supporto per nuovi chipset grafici, assicurando funzioni di sospensione e ripristino su un maggior numero di piattaforme. In particolare, è stato notevolmente migliorato il supporto per i chipset NVIDIA, soprattutto per la serie GeForce 8800.

3.4. Kernel Tickless

In precedenza, il kernel Linux interrompeva periodicamente ogni CPU presente in un sistema, ad una determinata frequenza — 100 Hz, 250 Hz o 1000 Hz, secondo la piattaforma. Il kernel interrogava la CPU sui processi in esecuzione, ed in base ai risultati stabiliva l'ordine di esecuzione dei processi ed equilibrava il carico sulla CPU. Conosciuto come *timer tick*, il kernel interrompeva la CPU senza alcuna considerazione sullo stato energetico della CPU. Una CPU inattiva poteva essere riattivata anche 1000 volte al secondo e su sistemi che implementavano misure di risparmio energetico per CPU inattive, il timer tick impediva di fatto di ricavare ogni beneficio al riguardo.

Il kernel in Fedora 14 esegue in modalità *tickless*, cioè, le precedenti interruzioni periodiche del timer sono state sostituite con interruzioni su-richiiesta. In tal modo, le CPU inattive restano tali, fino

all'arrivo di un nuovo task da elaborare, ed inoltre risulta prolungata la permanenza delle CPU negli stati a *minore energia*.

3.5. Active State Power Management

ASPM (Active State Power Management), è un protocollo per il risparmio energetico nei sottosistemi *PCI Express* o *PCIe* (Peripheral Component Interconnect Express), impostando nei link PCIe, uno stato a basso assorbimento di energia quando le periferiche connesse non sono in uso. ASPM controlla lo stato di assorbimento ad entrambe le estremità del link, riducendo il consumo di energia nel link anche quando la periferica si trova a pieno regime operativo.

Tuttavia, quando ASPM è abilitato, aumenta il tempo di risposta a causa della transizione richiesta dal link per passare tra i diversi stati di alimentazione. ASPM presenta tre stati di alimentazione:

default

imposta gli stati di alimentazione del link PCIe in base alle impostazioni predefinite specificate nel firmware di sistema (per esempio, BIOS). Questo è lo stato predefinito di ASPM.

powersave

imposta ASPM sul risparmio energetico appena possibile, prescindendo dalle prestazioni.

performance

disabilita ASPM, permettendo ai link PCIe di operare a prestazioni massime.

Le policy ASPM sono impostate in `/sys/module/pcie_aspm/parameters/policy`, e possono essere specificate anche all'avvio del sistema, usando il parametro del kernel `pcie_aspm`:

pcie_aspm=off e **pcie_aspm=force**, rispettivamente disabilitano ed abilitano ASPM anche su dispositivi senza supporto ASPM.

Avvertenza — `pcie_aspm=force` può causare il blocco del sistema

Impostando **pcie_aspm=force**, l'hardware privo di supporto ASPM può causare il blocco del sistema. Quindi, prima di impostare **pcie_aspm=force** assicurarsi che tutti gli hardware PCIe sul sistema supportino ASPM.

3.6. Aggressive Link Power Management

ALPM (Aggressive Link Power Management), è una tecnica di risparmio energetico impiegata nei dischi, impostando un livello energetico basso nel link SATA del disco, durante il periodo di inattività (cioè, in assenza di I/O). ALPM, in presenza di una coda di I/O su link, ripristina automaticamente il link SATA su uno stato attivo a più alta energia.

Tuttavia, il risparmio generato da ALPM avviene a spese di latenze nel disco, perciò si consiglia di utilizzare ALPM solo se si prevedono lunghi periodi con assenza di I/O.

ALPM è disponibile solo su controller SATA che utilizzano *AHCI* (Advanced Host Controller Interface). Per maggiori informazioni su AHCI, fare riferimento all'[AHCI Specification](#)¹.

Quando disponibile, ALPM è abilitato per impostazione. ALPM presenta tre modalità:

¹ <http://www.intel.com/technology/serialata/ahci.htm>

min_power

In assenza di I/O, imposta il link SATA sullo stato a energia minima (SLUMBER). Questa modalità può risultare utile quando si prevedono lunghi periodi di inattività.

medium_power

In assenza di I/O, imposta il link SATA sullo stato a energia intermedia (PARTIAL). Questa modalità è stata progettata per accelerare le transizioni tra stati (p.e. durante i periodi intermittenti tra I/O intensi e I/O ridotti), per ridurre al minimo l'impatto sulle prestazioni.

La modalità `medium_power` serve ad agevolare la transizione del link SATA, a seconda del carico di I/O, tra gli stati `PARTIAL` (a potenza media) e `ACTIVE` (a potenza massima). Da notare che non è possibile una transizione diretta da uno stato `PARTIAL` a `SLUMBER` e viceversa: gli stati non possono transitare da un stato all'altro, senza prima passare per lo stato `ACTIVE`.

max_performance

ALPM è disabilitato; in tal caso, in assenza di I/O, il link SATA non può entrare in uno stato a basso consumo.

Gli adattatori host SATA supportano ALPM, se esiste il file `/sys/class/scsi_host/host*/link_power_management_policy`. Per controllare le impostazioni correnti, vedere il file; per modificare le impostazioni, usare i parametri descritti in questa sezione.



Importante — alcune impostazioni disabilitano l'hot plug

Impostando ALPM su `min_power` o `medium_power` si disabilita automaticamente la funzionalità di "Hot Plug".

3.7. Ottimizzazione di accesso al disco Relatime

Lo standard POSIX richiede che i sistemi operativi mantengano metadati del file system, in cui salvare/registrare i dati di accesso ai file. Questo timestamp è detto `atime`, e la sua gestione richiede una serie costante di operazioni di scrittura sui dischi, mantenendoli occupati e attivi. Poiché poche applicazioni fanno uso di `atime`, questa attività del disco consuma energia. Addirittura, la scrittura si verifica anche se il file non è letto dal disco ma caricato dalla cache. Per un certo periodo, il kernel Linux ha supportato un'opzione `noatime` per `mount`, che non scriveva dati `atime` sui file system montati con questa opzione. Tuttavia, disabilitare in questo modo questa attività di `atime`, crea problemi a quelle applicazioni che usano i dati scritti da `atime`.

Il kernel usato in Fedora 14, supporta un'altra opzione — `relatime`. `relatime` conserva i dati `atime`, ma non per ogni accesso ai file. Abilitando questa opzione, i dati `atime` vengono scritti su disco solo se il file risulta modificato rispetto all'ultimo aggiornamento (`mtime`) dei dati di `atime`, o se l'ultimo accesso al file risale a un certo periodo di tempo (un giorno, per impostazione).

Per impostazione, tutti i file system sono montati con l'opzione `relatime`. Per annullare questa caratteristica su tutto il sistema, all'avvio del sistema passare il parametro di boot `default_relatime=0`. Se `relatime` è abilitato per impostazione su un sistema, è possibile comunque annullarlo su un file system, montando il file system con l'opzione `norelatime`. Infine, per modificare il periodo di tempo predefinito per l'aggiornamento dei dati `atime` di un file, usare il parametro di boot `relatime_interval=`, specificando il periodo in secondi. Il valore predefinito è **86400**.

3.8. Power Capping

Fedora 14 supporta le funzionalità di *Power Capping* dei recenti hardware, come le tecnologie *DPC* (Dynamic Power Capping) di HP e *Node Manager* (NM) di Intel. Il *Power Capping* permette agli amministratori di limitare i consumi nei server e di pianificare con più efficienza i data center, diminuendo i rischi di sovraccarico di potenza, soprattutto nei locali con un gran numero di server, dato che la potenza assorbita da ciascuno di essi non potrà eccedere il limite fissato.

HP Dynamic Power Capping

Dynamic Power Capping, funzione disponibile su server ProLiant e BladeSystem, permette agli amministratori di limitare il consumo energetico di un server o un gruppo di server. Si tratta di una limitazione definitiva che il server non può superare, a prescindere dal carico di lavoro. Questa funzione di limitazione ha effetto solo quando il server approssima l'assorbimento di potenza massima, a quel punto interviene un gestore dedicato che agendo sugli stati P ed il clock della CPU, riduce la potenza a livelli inferiori.

Dynamic Power Capping interviene sugli stati della CPU indipendentemente dal sistema operativo, comunque, il firmware *integrated Lights-Out 2* (iLO2) di HP, consente ai sistemi operativi di accedere al gestore del processore e quindi anche le applicazioni dello spazio utente possono interrogare il gestore del processore. Il kernel usato in Fedora 14 include un driver per firmware HP iLO e iLO2, con cui i programmi possono interrogare il gestore dei processori, tramite il file `/dev/hpilo/dXccbW`. Il kernel include anche una estensione di `hwmon` dell'interfaccia `sysfs` a supporto del *power capping*, e un driver di `hwmon` per misure di potenza in sistemi ACPI 4.0 che usano l'interfaccia `sysfs`. Tutte queste caratteristiche permettono al sistema operativo e ai tool dello spazio utente di leggere il limite configurato per `power cap` e la potenza assorbita dal sistema.

Per ulteriori dettagli, fare riferimento la sito ufficiale: [HP Dynamic Power Capping](#)².

Intel Node Manager

Node Manager di Intel, usando gli stati P e T della CPU, impone un *power cap* per limitare le prestazioni e quindi gli assorbimenti di potenza. Con una policy di gestione energetica adeguata, gli amministratori possono ridurre i consumi, adeguando i sistemi ai carichi di lavoro, per esempio riducendo gli assorbimenti di potenza nei periodi di basso carico, come durante la notte o i fine settimana.

Node Manager di Intel, regola le prestazioni della CPU con *OSPM* (Operating System-directed configuration and Power Management) dello standard *ACPI* (Advanced Configuration and Power Interface). Quando Node Manager notifica al driver *OSPM* le variazioni di stati T, il driver esegue le corrispondenti variazioni negli stati P del processore. In modo simile, quando Node Manager notifica al driver *OSPM* le variazioni di stati P, il driver esegue le corrispondenti variazioni negli stati T. Queste variazioni avvengono automaticamente e non richiedono alcun intervento da parte del sistema operativo. La configurazione e la gestione di Node Manager si avvale del programma *DCM* (Data Center Manager) di Intel.

Fare riferimento al documento [Node Manager — A Dynamic Approach To Managing Power In The Data Center](#)³, per maggiori informazioni.

² <http://h18013.www1.hp.com/products/servers/management/dynamic-power-capping/index.html>

³ <http://communities.intel.com/docs/DOC-4766>

3.9. Migliore Gestione Energetica nei Sistemi Grafici

Fedora 14 risparmia energia in monitor e periferiche video, eliminando varie fonti di consumo non indispensabili.

Reclocking LVDS

LVDS (Low-voltage differential signalling), è un sistema di trasmissione di segnali elettronici, usato soprattutto nei Notebook, per trasmettere informazioni sui pixel degli schermi *LCD* (Liquid Crystal Display). Tutti i display hanno una *frequenza d'aggiornamento* o refresh — la frequenza alla quale un controller grafico trasmette nuovi dati e viene ricostruita la schermata — generalmente di 60 Hz. Se il controller grafico ed il controller dello schermo sono collegati tramite LVDS, il sistema LVDS assorbe energia durante ogni ciclo di refresh. Durante uno stato inattivo, la frequenza di refresh di molti LCD può ridursi fino a 30 Hz, senza notabili effetti (diversamente dai monitor *CRT* (Cathode Ray Tube), dove una analogica riduzione produce un percepibile tremolio). Il driver delle schede grafiche Intel, inserito direttamente nel kernel in Fedora 14, effettua automaticamente questo *downclock* con un risparmio di circa 0.5 W, a schermo inattivo.

Abilitato l'Auto-Refresh nella SDRAM

Una memoria *SDRAM* (Synchronous Dynamic Random Access Memory), come quella usata nelle schede grafiche, viene *alimentata* migliaia di volte al secondo, per permettere alle celle interne di preservare i dati contenuti. Il controller della memoria, oltre a gestire il flusso dei dati, deve anche gestire questi cicli di refresh. Le SDRAM supportano anche una modalità di *auto refresh* a basso consumo. Questa modalità utilizza un timer interno che generando propri cicli di refresh, arresta il controller della memoria, senza perdita di dati. Il kernel usato in Fedora 14 attiva l'auto-refresh della memoria su schede grafiche Intel, con un risparmio di circa 0.8 W, quando inattive.

Riduzione del clock della GPU

Le GPU (Graphical Processing Units) tipiche, sono fornite di diversi clock che regolano le varie parti dei componenti interni al circuito. Il kernel usato in Fedora 14 può ridurre la frequenza di alcuni dei clock interni di unità GPU, ATI ed Intel, contribuendo ad un risparmio di energia. Il kernel riduce/aumenta automaticamente la velocità dei clock a secondo dell'attività della GPU, con un risparmio di potenza che arriva fino a 5 W.

Disattivazione della GPU

In Fedora 14 i driver grafici di ATI e di Intel riescono a rilevare il collegamento o meno di un monitor, disattivando completamente la GPU quando nessun monitor è collegato. Questa caratteristica si rivela particolarmente significativa nei server, ai quali di solito non viene collegato nessun monitor.

3.10. RFKill

Molti computer sono forniti di radiotrasmittitori, tra cui sistemi di trasmissione 3G, Wi-Fi e Bluetooth. Questi dispositivi assorbono potenza, che in genere viene sprecata quando non utilizzati.

RFKill è un sottosistema del kernel Linux che fornisce un'interfaccia per attivare/disattivare i radiotrasmittitori del sistema. Quando vengono disattivati, i trasmettitori possono essere messi in uno stato in cui i software possono riattivarli (un *soft block*) o non riattivarli (un *hard block*).

Il nucleo centrale di RFKill, fornisce le API (Application Programming Interface) del sottosistema. I driver del kernel sviluppati a supporto di RFKill, usano questa API per registrarsi nel kernel ed includono metodi per abilitare e disabilitare il dispositivo. Inoltre, il nucleo di RFKill fornisce metodi

per attivare notifiche che le applicazioni possono interpretare, e metodi per interrogare gli stati del trasmettitore.

L'interfaccia di RFKill si trova in `/dev/rfkill`, contenente lo stato corrente di tutti i radiotrasmettitori nel sistema. Ogni dispositivo registra il proprio stato in `sysfs`. Inoltre RFKill emette `uevents` ad ogni variazione di stato nei dispositivi controllati da RFKill.

Rfkill è uno strumento da terminale, con il quale interrogare e modificare i dispositivi riconosciuti da RFKill. **Rfkill** si trova nel pacchetto `rfkill`.

Con il comando `rfkill list`, si ottiene un elenco di dispositivi, ognuno dei quali ha un *index number* associato, a partire da `0`. Usare l'*index number* per indicare a `rfkill` il dispositivo da bloccare o sbloccare, per esempio:

```
rfkill block 0
```

blocca il primo dispositivo riconosciuto da RFKill.

Si possono bloccare anche certe categorie di dispositivi o tutti i dispositivi. Per esempio:

```
rfkill block wifi
```

blocca tutti i dispositivi Wi-Fi presenti nel sistema. Per bloccare invece tutti i dispositivi, eseguire:

```
rfkill block all
```

Per sbloccare i dispositivi, eseguire `rfkill unblock`. Per ottenere l'elenco delle categorie di dispositivo che `rfkill` è in grado di bloccare, eseguire `rfkill help`.

3.11. Ottimizzazione nello Spazio Utente

Ridurre l'attività svolta dall'hardware contribuisce a ridurre i consumi. Nel [Capitolo 3, Infrastruttura Centrale e Meccanismi](#), sono state descritte le configurazioni che permettono al sistema di operare in vari stati a basso consumo, tuttavia le applicazioni dello spazio utente sembrano annullare gli sforzi fatti, richiedendo all'hardware operazioni spesso inutili ed impedendo di passare in questi stati ad energia minima. Durante lo sviluppo di Fedora 14, diversi audit sono stati eseguiti nelle seguenti aree, con l'obiettivo di ridurre le operazioni hardware, non indispensabili:

Riduzione dei wakeup

Fedora 14 usa il kernel *tickless* (vedere la [Sezione 3.4, «Kernel Tickless»](#)), permettendo alla CPU una permanenza più lunga, negli stati a bassa/ridotta attività. Comunque, non solo il *timer tick* nei precedenti kernel, è l'unica causa che genera eccessivi wakeup della CPU, anche le chiamate di funzione, nelle applicazioni, contribuiscono in buona parte. In Fedora 14 le chiamate di funzione non necessarie, sono state ridotte in oltre 50 applicazioni.

Riduzione delle operazioni di I/O

Anche le operazioni di I/O (Input/Output) dei dischi e delle periferiche di rete, consumano energia. Nelle periferiche di rete e nei dischi, in cui sono disponibili stati a minima energia (p.e. ALPM o ASPM), quando inattive, un flusso di dati non indispensabile, di fatto, impedisce ai dispositivi di passare e permanere in questi stati, impedendo, nel caso dei dischi di ridurre persino la velocità di rotazione. Richieste eccessive e non necessarie, sono state minimizzate in numerose applicazioni, in particolare quelle che impedivano di rallentare la rotazione dei dischi.

Audit negli script di Init

I servizi avviati automaticamente, richiesti o meno che siano, presentano un grande potenziale a consumare risorse. Quindi i servizi dovrebbero essere impostati, quando possibile, su "off" o su "on demand". Per esempio, il servizio **BlueZ** che abilita il supporto per Bluetooth, in precedenza veniva eseguito automaticamente all'avvio del sistema, senza tener conto della presenza o meno di hardware Bluetooth. Ora invece, prima di avviare il servizio, lo script di init di **BlueZ**, controlla la presenza di hardware Bluetooth.



DRAFT

Casi d'Uso

Questo capitolo descrive due casi d'uso, illustrando i metodi di analisi e di configurazione, descritti in altre sezioni di questa guida. Nel primo caso vengono presi in considerazione server tipici, nel secondo un portatile.

4.1. Caso Server

Oggi giorno, un tipico server standard, viene commercializzato sostanzialmente con tutto l'HW indispensabile, supportato in Fedora 14. La prima cosa da considerare è il carico di lavoro che il server dovrà gestire. In base a queste informazioni sarà possibile decidere i componenti da ottimizzare per il risparmio energetico.

A prescindere dal tipo di server, generalmente non si richiedono prestazioni grafiche, perciò si può abilitare il risparmio energetico per l'unità GPU.

Webserver

Un webserver ha bisogno di una rete e di un disco. In base alla velocità del collegamento esterno, 100 Mbit/s dovrebbero bastare. Se la macchina serve soprattutto pagine statiche, le prestazioni della CPU non dovrebbero essere molto importanti. Le scelte volte al risparmio energetico, potrebbero quindi includere:

- nessun plugin di regolazione di **tuned** per rete o disco.
- ALPM abilitato.
- regolatore ondemand abilitato.
- scheda di rete limitata a 100 Mbit/s.

Server di elaborazione

Un server di elaborazione necessita di CPU. Le scelte per il risparmio energetico, potrebbero includere:

- secondo l'attività e il tipo di archiviazione dei dati, plugin di regolazione di **tuned** per disco o rete; o per sistemi batch-mode, **tuned** completamente attivo.
- secondo il tipo di utilizzo, volendo anche il regolatore per performance.

Mailserver

Un server di posta ha bisogno soprattutto di I/O disco e CPU. Le scelte per il risparmio energetico, potrebbero includere:

- regolatore ondemand attivato (le prestazioni della CPU nelle ultime cifre percentuali sono irrilevanti).
- nessun plugin di regolazione di **tuned** per rete o disco.
- la velocità di rete non dovrebbe essere limitata, poichè il servizio di posta molto spesso è interno alla rete, quindi dovrebbe beneficiare di un link a 1 Gbit/s o 10 Gbit/s.

Fileserver

I requisiti per il fileserver sono simili a quelli del server di posta, ma dipendendo dal protocollo usato, potrebbe essere necessaria una CPU dalle prestazioni più elevate. Generalmente i server basati su

Samba richiedono più CPU dei server NFS, ed i server NFS più CPU dei server iSCSI. In ogni caso, si dovrebbe usare il regolatore ondemand.

Directory server

Un directory server, generalmente presenta bassa attività di I/O disco, soprattutto se equipaggiato con RAM adeguata. La latenza di rete è importante anche se l'I/O di rete è minima.

4.2. Caso Laptop

Un altro insieme di sistemi, in cui la gestione energetica ed il conseguente risparmio possono realmente fare la differenza, sono i portatili. Poiché già durante tutta la fase di realizzazione, i portatili utilizzano una minore quantità di energia rispetto alla realizzazione di una workstation o server, in termini assoluti, il potenziale per risparmi ulteriori risulta inferiore rispetto alle altre macchine. Ad ogni modo, conoscere ed applicare qualche utile accorgimento di tipo energetico, quando il portatile è alimentato a batteria, può certamente contribuire a prolungare la vita media della batteria. Anche se questa sezione è dedicata soprattutto all'uso dei portatili in modalità batteria, grosso modo le regolazioni saranno applicabili anche usando l'alimentazione di rete AC.

Il risparmio energetico su singoli componenti, generalmente apporta un beneficio relativamente maggiore nei portatili che nelle workstation. Per esempio, una scheda di rete da 1 Gbit/s che funziona a 100 Mbits/s risparmia dai 3 ai 4 Watt. In un server tipico, con un consumo totale di circa 400 Watt, tale risparmio rappresenta circa l'1 %. In un portatile con un consumo totale di circa 40 Watt, il risparmio energetico ammonta al 10 % del totale.

Specifiche ottimizzazioni di consumo che è possibile applicare ai portatili, includono:

- Disabilitare nel BIOS, tutto l'hardware non utilizzato. Per esempio, porte parallele/seriali, lettori di smart-card, webcam, WiFi e Bluetooth.
- Ridurre la luminosità dello schermo in ambienti poco illuminati, per una lettura più confortevole. Usare **Sistema+Preferenze** → **Gestione alimentazione** nel desktop GNOME, oppure **Kickoff Application Launcher+Computer+Impostazioni sistema+Avanzate** → **Gestione alimentazione** nel desktop KDE, oppure avviare la GUI con i comandi **gnome-power-manager** o **xbacklight**, o usare i tasti funzione sul portatile.
- Usare il profilo `laptop-battery-powersave` del regolatore **tuned-adm**, per abilitare i relativi meccanismi di risparmio energetico. Nota: *Ciò influenza le prestazioni e la latenza della scheda di rete e del disco.*

In aggiunta (o alternativamente) si possono apportare piccole regolazioni a varie impostazioni di sistema:

- usare il regolatore ondemand (abilitato per default in Fedora 14)
- abilitare la modalità laptop (parte del profilo `laptop-battery-powersave`):

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

- aumentare il flush-time del buffer del disco (parte del profilo `laptop-battery-powersave`):

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

- disabilitare il watchdog su [nmi \(non-maskable interrupt\)](#)¹ (parte del profilo `laptop-battery-powersave`):


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

- abilitare il risparmio energetico per il processore audio AC97 (abilitato per impostazione, in Fedora 14):

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

- abilitare il risparmio energetico per i processori multi-core (parte del profilo laptop-battery-powersave):

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

- abilitare l'auto-sospensione dei sistemi USB:

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

Nota: *l'auto-sospensione non funziona correttamente con tutti i dispositivi USB.*

- abilitare gli stati a minima energia per ALPM (parte del profilo laptop-battery-powersave):

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

- montare i filesystem con l'opzione relatime (per impostazione, in Fedora 14):

```
mount -o remount,relatime mountpoint
```

- attivare per i dischi, la migliore modalità di risparmio energetico (parte del profilo laptop-battery-powersave):

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

- disabilitare il polling del drive CD/DVD (parte del profilo laptop-battery-powersave):

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

- ridurre la luminosità dello schermo a **50** o meno, per esempio:

```
xbacklight -set 50
```

- attivare il DPMS per schermo inattivo:

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

- ridurre il livello di emissione del sistema Wi-Fi (parte del profilo laptop-battery-powersave):

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

- disattivare il Wi-Fi:

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

- limitare la rete ethernet a 100 Mbit/s (parte del profilo laptop-battery-powersave):

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

DRAFT

Appendice A. Suggerimenti per Sviluppatori

Ogni buon manuale di programmazione affronta i problemi di allocazione memoria e le prestazioni di specifiche funzioni, un po' meno le questioni legate al risparmio energetico. Come buona pratica di sviluppo del software, quindi, si dovrebbero considerare anche quelle situazioni che possono influire sui consumi energetici nei sistemi sui quali il software dovrà funzionare. Queste considerazioni non influenzano ogni riga di codice, ma sono dirette ad ottimizzare quelle parti di codice che più frequentemente presentano limitazioni nelle prestazioni.

Alcune tecniche che risultano spesso problematiche, sono:

- utilizzo di thread.
- wakeup non necessari della CPU unito al loro uso poco efficiente. Alternativa: se occorre un wakeup, *race to idle*, eseguire i task il più velocemente possibile.
- utilizzo non indispensabile della funzione `[f] sync()`.
- utilizzo non necessario di polling o di brevi regolari timeout. Alternativa: reagire agli eventi.
- utilizzo non efficiente dei wakeup.
- accesso inefficiente al disco. Alternativa: usare buffer capienti per ridurre gli accessi; usare blocchi di scrittura più grandi.
- utilizzo inefficiente dei timer. Alternativa: raggruppare, quando possibile, i timer di più applicazioni, o meglio ancora dei sistemi.
- utilizzo eccessivo dell'I/O e della memoria, inclusi i *memory leaks*¹
- esecuzioni non necessarie di elaborazioni

Le seguenti sezioni esaminano in maggior dettaglio, alcune di queste aree.

A.1. Usare i Thread

È un luogo comune credere che l'uso dei thread migliori le prestazioni e la velocità delle applicazioni: ciò non è sempre vero!.

Python

Python usa GIL o Global Lock Interpreter², che rende vantaggioso il threading solo in significative operazioni di I/O. **unladen-swallow**³ è una implementazione di Python più veloce, con cui è possibile ottimizzare il codice.

Perl

I thread di Perl sono stati originariamente creati per eseguire applicazioni senza usare processi di fork (come nei sistemi windows a 32-bit). Nei thread di Perl, i dati vengono copiati per ogni singolo thread (Copy On Write). Per impostazione, i dati non vengono condivisi e gli utenti dovrebbero definire il loro

² <http://docs.python.org/c-api/init.html#thread-state-and-the-global-interpreter-lock>

³ <http://code.google.com/p/unladen-swallow/>

livello di condivisione. Per la condivisione dei dati, occorre includere il modulo `threads::shared`. In tal modo, il modulo non solo copia i dati (Copy On Write), ma crea anche delle variabili ad essi collegati, con l'effetto di richiedere più tempo e di risultare più lento.⁴

C

I thread di C condividono la stessa memoria, ogni thread ha il proprio stack ed il kernel non deve creare nuovi descrittori di file ed assegnare nuova memoria. Il C può effettivamente avvalersi del supporto di più CPU per più thread. Perciò, per massimizzare le prestazioni dei thread, usare un linguaggio low-level come C o C++. Se si usa un linguaggio di script, si consideri la progettazione di un *C binding*. Impiegare un software di *profiling* per identificare nel codice, le sezioni con scarse prestazioni.⁵

A.2. Wakeup

Molte applicazioni scansionano i file di configurazione alla ricerca di variazioni. In molti casi, la scansione avviene ad intervalli regolari, per esempio ogni minuto. Ciò può essere un problema, perchè forza la riattivazione del disco dal suo stato di spindown (lo stato inattivo, in cui il disco ruota a minime velocità). La soluzione migliore sarebbe impostare un intervallo adeguato o un buon meccanismo di controllo, oppure verificare le variazioni con **inotify** e reagire agli eventi. **inotify** può controllare un gran numero di variazioni sia in file sia in directory.

Per esempio:

```
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();
}
...
```

Il vantaggio di questo approccio è dato dalla varietà di controlli che è possibile effettuare.

Il principale svantaggio è rappresentato dal limitato numero di controlli disponibili nel sistema, indicato nel file `/proc/sys/fs/inotify/max_user_watches` che si raccomanda di non modificare. Inoltre, nel caso in cui **inotify** fallisca, il codice dovrebbe ricorrere ad un metodo di controllo alternativo; il che significa inserire nel codice molte istruzioni `#if #define`.

⁴ http://www.perlmonks.org/?node_id=288022

⁵ <http://people.redhat.com/drepper/lt2009.pdf>

Per maggiori informazioni su **inotify**, consultare la pagina man su `inotify`.

A.3. fsync

`fsync` è nota essere una operazione di I/O molto esigente, ma ciò non è completamente vero. Per spezzare una lancia a favore di `fsync`, consultare l'articolo scritto da Theodore Ts'o' *Don't fear the fsync!*⁶ e la relativa discussione.

Firefox usava caricare la libreria **sqlite** ogni volta che l'utente cliccava sul link di una nuova pagina. **sqlite** a sua volta chiamava la funzione `fsync` e a causa delle impostazioni del file system (soprattutto del f.s. `ext3` in modalità `data-ordered`), si verificava una lunga latenza. La pagina web si caricava dopo un lungo ritardo, che poteva durare fino a 30 secondi se era in esecuzione un altro processo, che per esempio, copiava un file molto grande.

Inoltre, in altri casi, in cui `fsync` non veniva per nulla usata, si verificarono problemi con il passaggio al file system `ext4`. `ext3` era impostato in modalità `data-ordered`, che ripuliva la memoria ogni pochi secondi, salvando i dati sul disco. Ma con `ext4` e `laptop_mode` (usato nei portatili per prolungare la durata della batteria), l'intervallo delle operazioni di salvataggio era più lungo e poteva esserci una potenziale perdita di dati se il sistema veniva spento, improvvisamente. Ora `ext4` è stato *patched*: (morale della favola) progettare attentamente le applicazioni e usare in modo appropriato la funzione `fsync`.

Il seguente semplice esempio di lettura e scrittura in un file di configurazione, mostra come sia possibile eseguire un backup di un file o perdere i dati:

```
/* 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);
```

Questa sarebbe una soluzione migliore:

```
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");
```

⁶ <http://think.org/tytso/blog/2009/03/15/dont-fear-the-fsync/>

DRAFT

Appendice B. Cronologia Revisione

Revisione 0.1 Thu Jul 29 2010

Rüdiger Landmann
r.landmann@redhat.com

Importazione in Upstream



DRAFT