

第97回カーネル読書会 TOMOYO Linuxメインライン化記念勉強会

NILFSのメインライン化 について

2009年7月3日

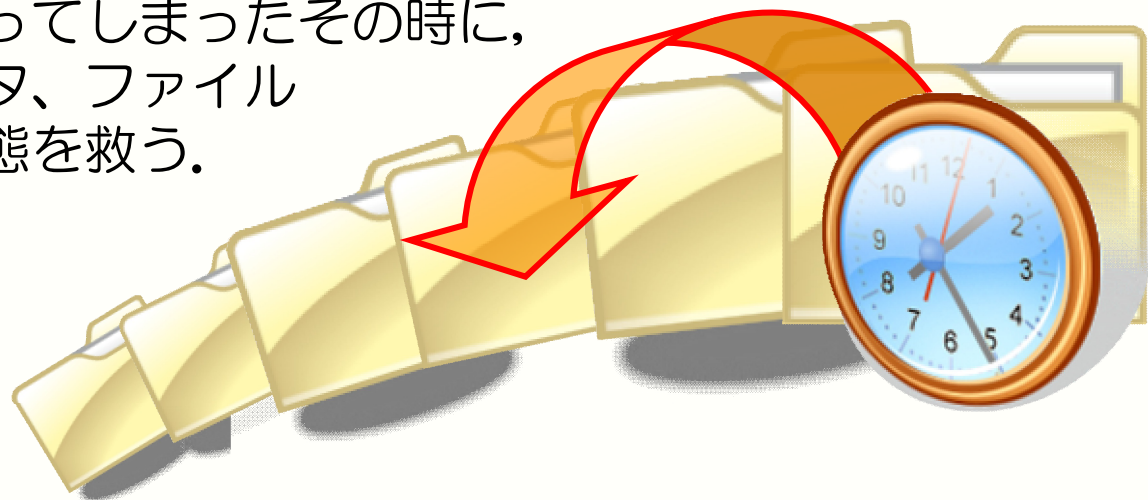
Signed-off-by: Ryusuke Konishi
<konishi.ryusuke@lab.ntt.co.jp>

NILFSって何だっけ？

- NILFSは「連続スナップショット」機能を持つLinux向けの ログ構造化ファイルシステム

NILFS = A **N**ew **I**mplementation of the
Log-structured **F**ile **S**ystem

- ▶ ファイルシステムの過去の状態をまるごと履歴として残せる
- ▶ 誤って上書き・削除しても、直前の状態を取り戻せる。
 - 「あっ」と言ってしまったその時に、ユーザのデータ、ファイルシステムの状態を救う。



なぜ NILFS?

2007年4月7日 朝日新聞朝刊(14版) 39面

ヤフーのウェブメールサービス「Yahoo!メール」で、会員約28万人の受信メール約450万通(昨年12月〜今年2月分)の本文部分を誤って消し、回復できなくなった。6日の同社の発表によると、ウェブメールのデータは個人のパソコンでなく運営者のサーバーに保存されるが、このサーバーのプログラムミスで、迷惑メールを消す処理が一般メールに適用されたためだという。会員は約1300万人。99年1月のサービス開始以来、データを誤って消去したのは初めてと

メール450万通消す ヤフー、サーバー設定ミス

いう。他の会員約40万人の約515万通のメールも一時的に読めなくなつたが、残ったデータから回復させた。

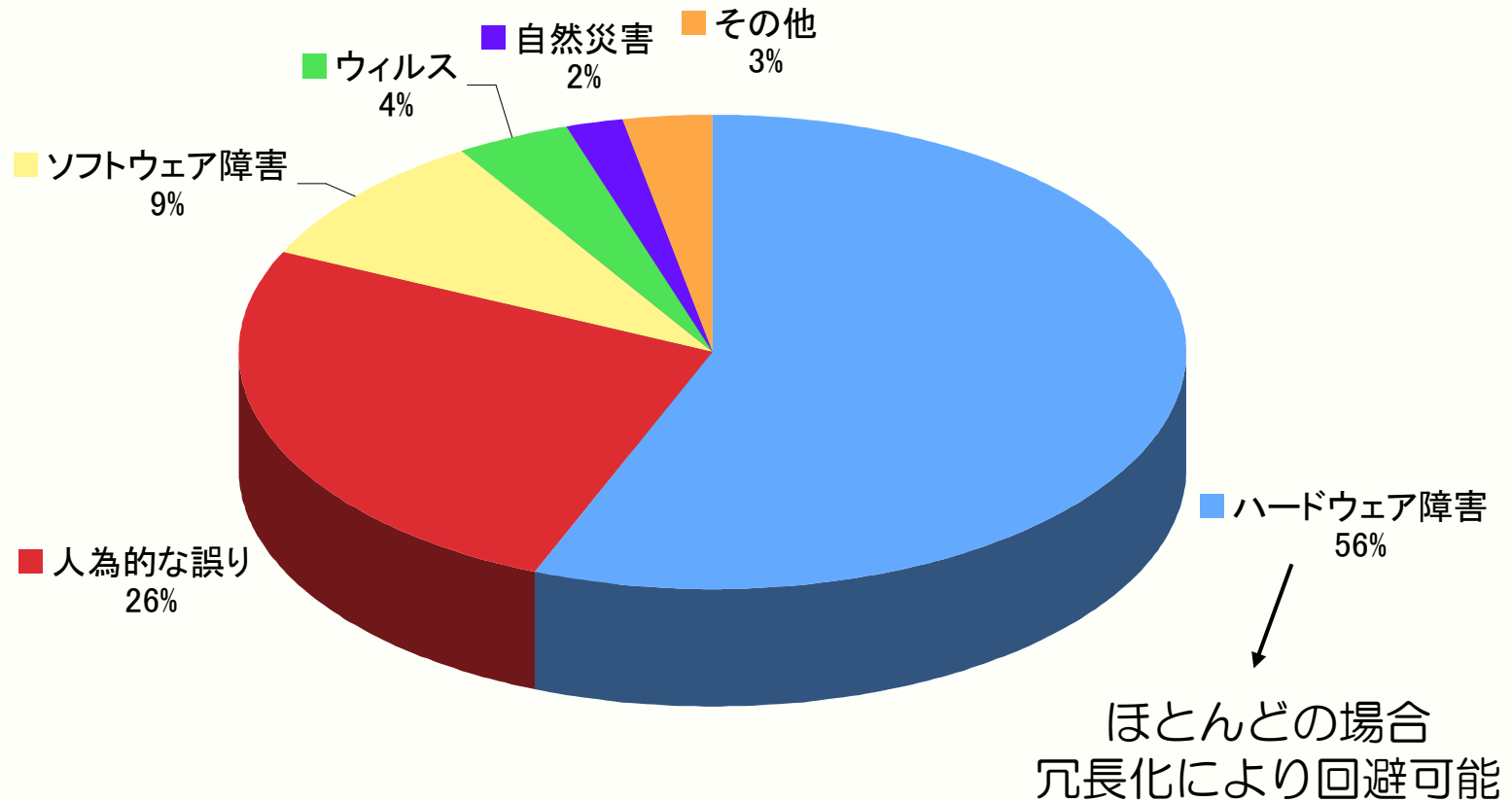
設定のミスにより
450万通のメールが復旧不能

バックアップから
515万通のメールを復旧

約1000万通を
誤って消去

- ① 重大なシステム障害は人間のミス (バグ, 設定ミス, 操作ミス) で起きる
- ② バックアップでの100%復旧は困難

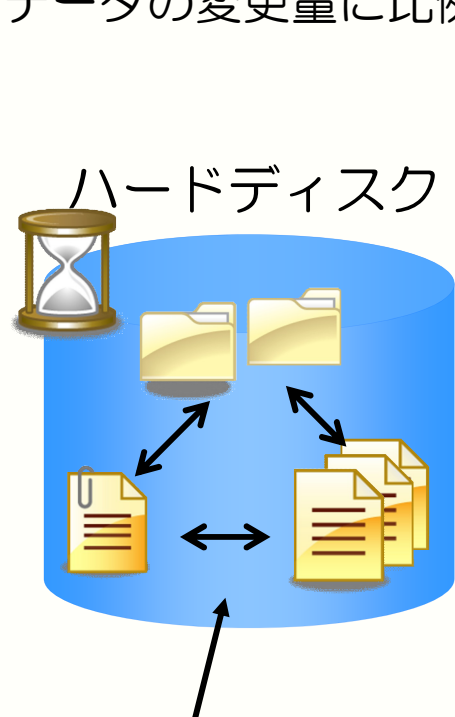
データ喪失の原因



ハードディスク等からのデータリカバリの専門の Kroll Ontrack 社の調査による。オフィスPCを含む。

バックアップの問題

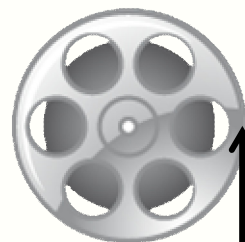
取得に時間がかかる
データの変更量に比例



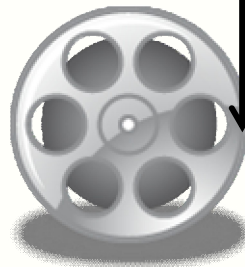
一貫した状態をとる
には作業中断、停止
が必要



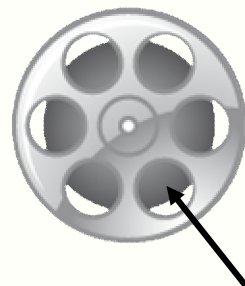
バックアップメディア



取得間隔が離散的
8時間毎, 1日毎等



最終バックアップ後の
データ・変更は保護さ
れない



直接データにアクセスできない
リストアが必要

メインライン化の道のり(1)

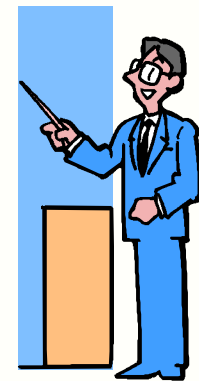
■ 構想は2003年秋ごろから

- ▶ スナップショット機能を実装した高信頼なファイルシステムをLinuxで実現しよう!
- ▶ ログ構造化ファイルシステムはどうだろう
- ▶ ーからファイルシステムを書くのは大変, LinLogFS を改良・発展させよう



メインライン化の道のり(2)

- 2004年12月 NILFS開発始動
 - ▶ B-tree採用の新規LFS
- 2005年9月 NILFS(v1) OSS公開
- 2006年2月 YLUGカーネル読書会にてNILFSを紹介(天海)
- 2006年6月 OSDL の BOF で提案
 - ▶ Andrew Morton氏からGC (ごみ集め機能)ができれば mmツリーに入れて良いと言われる



メインライン化の道のり(3)

- ゴミ集め機能(GC)はv1でついに実現せず。設計をやり直すことに
- 2007年6月 NILFS2テスト版公開
- 2008年2月 NILFS2正式版公開。メインライン化作業本格化
 - ▶ 大幅な書き直し。2万8千行のコードは2万行に
- 2008年8月パッチ投稿, 9月に mm ツリーにマージされる。(だがレビューコメントが集まらない)
- 2008年12月頃, ioctlの設計に数々の問題発覚。不具合報告続出。しかしこれらが転機に

メインライン化の道のり(4)

- レビューコメントと不具合対応に追われる日々。いつのまにかメンテしている雰囲気か？
- 2009年3月 マージリクエスト。Andrew は **Yes** と言ってくれた!
- 息をひそめて成り行きを見守る。2009年4月 **2.6.30-rc1** マージ、同6月**2.6.30**リリース。



- ▶ 意に反し、海外メディアでSSD性能がクローズアップ
Linux magazine: “NILFS: A File System to Make SSDs Scream”
- ▶ 実は性能チューニングは後回し。評価したSSDでたまたま良かっただけです。

NILFS開発の反省

- ファイルシステムでよかったのか？
 - ▶ カーネルの中では比較的独立した機能。「マージは難しくない」？
 - ▶ 設計は楽しい。きちんと動くまでの遠い道のり。
 - ▶ ちょっとした不注意で、ユーザのデータが壊れる。
- 最初から開発コミュニティと連携すべきだった
 - ▶ 「作ってからマージして」の悪い例
 - ▶ コミュニティへの Respect 不足による多大な手戻り。

NILFSのこれから

- カーネル機能の継続的な改良
 - ▶ バグフィックス, 性能改善
- ユーザからの要望
 - ▶ ロールバック, 増分バックアップのサポート
 - ▶ GC(ごみ集め機能)の改良
 - ▶ SSD対応の強化
- よりオープンな開発体制への移行
- 対応アプリケーション・サービスの開発

TOMOYOおめでとう!

(NILFSもよろしく)

