

全学技術センター Linux サーバ障害

○大川敏生^{A)}、池田将典^{A)}

^{A)} 共通基盤技術支援室 情報通信技術系

概要

11月21日の東山地区の停電後に発生した、全学技術センターのLinuxサーバの障害について報告する。

当サーバは、ホームページ、メールサーバ等を運用しているため、システムを堅牢にするために、ハードウェア RAID5 を導入していた。システムの障害において、絶対安全という保証を得るには、高額なシステムの導入や、常時システムの状態をモニターするような運用体制が必要であるが、当サーバにおいては常時すぐ近くに担当者が居ないなど、十分な対応を行う事が出来なかった。

特に今回の障害では、RAID5 システムを過信した事に起因する人為的な点と、RAID コントローラの障害が原因であった。システムの早急な復旧を行うための方法を模索したが、元システムの修理ではなく、仮想サーバ上に同一環境を構築する方法を行った。

当サーバの、障害再発防止のために、皆様の日頃のシステム運用のノウハウについてご教示いただきますようお願いいたします。

1 当 Linux サーバの概要

当サーバは、2004年度に導入されたシステムである。全学技術センターの情報公開場として、ホームページ、電子メールの提供を行っている。ホームページでは、各種イベント、組織の紹介、業務依頼の受付等を行い、電子メールでは、メールの転送、メーリングリストによる情報窓口として利用されている。

また、当サーバは共通教育棟 II 4 階、全学技術センターサーバ室に設置され、技術職員が常駐しない場所で稼働しているため、可能な限り無人でも運用できるシステムを目標としている。

1.1 ディストリビュータ Vine3.2 の Linux システムによる運用

- ・ RAID5 構成によるハードディスクの障害対策
- ・ UPS による自動給電による停電対策

前述のとおり、定期的な保守があまり必要でなく、最近ではディストリビュータからのアップデートが無い状態であった。

1.2 バックアップデータについて

当サーバの RAID5 システムの堅牢であると過信した結果、システムの大幅なシステム変更時以外、バックアップを取っていなかった。保有する最新のバックアップは 2009 年の Tape メディアのみである。

1.3 新サーバの移行準備

連続運転にともなう消耗と思われる障害が発生しており、ハードディスクのエラーによるシステムの不調が発生し、ハードディスクの交換、RAID システムの自己修復が発生した。

一昨年度末に導入された新サーバ上に現サーバを移植する作業を試みる作業を行っていたが、デバイスドライバー等の不一致により、同一ディストリビュータ、同一バージョンの移植を断念し、Debian ディストリビュータの導入を行い、移植試験および、一部サービスの仮運用を行っていた。

移行のための作業時間がとれずに、そのままの状態が続き、仮想サーバ運用への転機が今年度の名大研修で訪れた。情報系内での勉強会も提案し、サーバ移行を皆で研鑽しながら行う予定をしていた時のシステム障害であった。

2 停電にともなう作業

数年前には、UPS システムによる停電対策を過信し、現場にも立ち会わずに定期停電を過ごしていた。実際、期待通りに自動シャットダウンから、復電時の自動起動までが行われていた。

昨年からは、先に述べた通りハードディスク等の不調等が発生しており、念のため停電時と復電時に立ち会うようにしていた。

2.1 事前、当日の作業

- ・ 前日に、停電当日の 30 分前に、システム自動シャットダウンの設定
- ・ 停電時の立ちたい
- ・ 自動シャットダウンの動作を目視で確認し、UPS のアラームを消す。
- ・ 復電後にシステムの動作確認
- ・ 復電時、システムの BIOS 起動画面に、RAID カードの警告が表示
- ・ Vine Linux の起動シーケンス開始
- ・ Fscck によるシステムチェック時にエラー発生
- ・ システム再起動により、BIOS 起動画面から、RAID カードの警告の解消を試みるが、回避できずに fsck のエラー発生までを数回繰り返す。
- ・ 1 人での作業が無理と判断
- ・ 障害を箕浦課長に報告し、帰宅

以上が障害発生までの経緯である。

3 復旧作業

サーバの修復にあたり、1 人作業での誤操作を防ぐために、池田副技師とともに修復作業を開始した。また、サーバの稼働を急務とし、新サーバによる Vine Linux の構築を開始した。Vine Linux の稼働については、稼働中のホームページサービスで利用されている PHP-Web サーバのバージョンに留意したためである。

3.1 サーバ、OS の復旧

- ・ RAID システムの修復作業

BIOS 上の RAID ユーティリティによる修復^[1]を数回試みるが、RAID の構成は復旧するが、データへのアクセスが不能となった。念のため、RAID 復旧業者^[2]へ連絡し電話による見積では、100 万円単位の費用がかかる旨であった。この時点で復旧作業を中断した。

- ・ 新サーバ上での再構築

先にも述べた通り、新サーバへの Vine Linux の構築が困難であるため、CentOS 上で Xen による仮想サーバを構築する方法を採用した。現行のサーバでは、概ね 10GB の領域で運用されていたため、仮想サーバにおいては 30GB の仮想サーバを構築した。

3.2 設定、コンテンツの復旧

- ・ システム移行作業時に使用した設定メモ

Vine Linux 移植の試行時に使用した現行のサーバ情報の設定情報のプリントアウトが煩雑な設定調節作

業を大幅に短縮することができた。現行のサーバではシステム、メール、ユーザ、ホームページ等各種サービス毎の担当者による運用が行われていたため、担当者からの設定情報を元にデータを再構築した。

- ・ 外部サービス^[3]によるコンテンツの取得

ホームページの情報については、外部サービスに保管されていた情報をもとに、古い情報ではあるがページの変遷程度の修復に留まった。

4 今後の対策

今回の障害の対策として、機器等による対策と運用による対策の2通りについて考察する。また、障害原因の発生頻度とその影響、事前対策の程度についても合わせて考察する。

4.1 機器等による対策

障害の原因と発生頻度を考慮した場合、電源、ネットワーク、サーバ本体の3系統に分けて対応する。本来であるならば、設置環境等を考慮し、当サーバの運用規模からトータルにシステムの構成を検討する必要があるが、費用や人的資源を考慮した場合、この最低限な対応となる。

- ・ 電源について

電源においては、UPSと協調した安全なサーバ稼働が現実でありである。現システムにおいても、UPSの電源状態の監視による安定電源の供給と、非常時のシステム自動停止、停止するまでの電源容量の確保を実現している。

- ・ ネットワークについて

当共通教育棟IIにおいて提供されているNICEネットワークのみである。ただし、配線等における障害を回避するためのHUB装置へのUPS電源供給と比較的安価なHUB本体の予備機を準備している。また、ネットワーク配線において、ラック等による配線設備が望ましい。

- ・ サーバについて

サーバの障害対策においては、サーバのデータ保持と、ホームページ、電子メールサービス等の対外的なサービス提供の保持を考慮するが、対外的なサービス提供を保持する場合、サーバ自身の予備機が必要と考える。対外的なサービスを保持する場合、ネットワークの多重化も考慮する必要があり、当センターの規模には過分と考える。したがって、データ保持を重視し、かつ代替え機等への移行が容易なシステムが望ましい。

今回の障害では、当時は高価であったハードディスクを有効に活用するためのRAID5構成が仇となってしまった。可能な限り単純で堅牢なシステムとして、RAID構成はミラーリングまでとし、サーバとは別の筐体のハードディスク等にバックアップを構築する構成が望ましい。

4.2 運用による対策

堅牢なシステムを保持するためには、運用による体制が不可欠である。理想として、より簡潔で明瞭な方法で、継続できる運用体制が望ましい。簡潔明瞭にすることで、特定の担当者でなくてもシステムの安定運用が期待できる。

- ・ システムの定期保守

当障害の初見で現れたサーバのエラーであるが、これらの対策として、定期的なfsckを行うことで回避できる障害である。ただし、サーバの再起動をとまなう可能性があるため、周知されたスケジュールにおいて手動により目視を含めた方法が望ましい。

- ・ 人的資源の効果的な体制

1 人管理システムは危険である。サービスの分担管理等で複数の人員で対応するシステムが望まれる。また、コンソール作業を容易にするために、遠隔地からのリモート作業を容易にすることで、保守性の向上を期待できる。

- ・ 仮想サーバの利点

最近普及している技術であるが、システムの汎用性を高くすることが可能となる。新サーバで導入した Xen 仮想サービスは、仮想マシン Vine Linux を 1 ファイルとして扱うことで、Xen の稼働する他のハードウェア上でもサーバを起動することが可能となる。つまり、ハードウェアに依存しないシステムとなり、機器の故障時においても、サービスの安定化が期待できる。また、1 ファイルの操作で、システム全体のバックアップが容易になり、保守性の向上が期待できる。

- ・ チェックシート・手順書

定期保守等の際に、前回と違う挙動を見つける方法が必要である。起動画面の情報は初期の障害対策に有用な情報を多く含むことがある。しかし、実際のコンソール画面上の情報は一部記録されても、流れてしまうだけで、人間の視力では前回の情報と比較することは不可能である。初代のメインフレーム等では、コンソールがラインプリンターであったりしたが、コンソールのログをとる仕組みの導入も望ましい。今回の障害時においても、携帯カメラのコンソール撮影が効果的であった。

5 さいごに

今回のサーバ障害におきまして、非常に多くの方から支援ならびに貴重な情報を頂きました。この場を借りてお礼申し上げます。本当にありがとうございました。本来あってはならないこのような事例に対処できたのは、人的繋がりがあったからこそと思います。あらためてお礼申し上げます。今後ともよろしく願いいたします。

参考文献

- [1] Adaptec Support Knowledgebase (よくある質問) ”故障したアレイは復旧できますか?”
(http://ask.adaptec.com/scripts/jp_adaptec_itic.cfg/php.exe/enduser/std_adp.php?p_faqid=17027)
- [2] RAID 修復業者 ”Ontrack Data Recovery Services” (<http://www.ontrack-japan.com/>)
- [3] Internet Archive: Digital Library of Free Books, Movires, Music & Wayback Machine “Wayback Machine”
(<http://www.archive.org/index.php>)