ディペンダブルITシステム構築技術

東 健二・上窪 真一・安場 洋輝

要　旨

ブロードバンドや携帯電話の普及により、ITシステムの利用度が高まるに伴い、その障害が引き起こす影響も大きくなっています。使いたい時に、使いたいサービスが、期待したレベルで使えるITシステム、すなわちディペンダブルなITシステムの構築には、高可用性の考え方が重要です。ハードウェアの経年劣化やOS、ミドルウェアなど複雑かつ多彩なソフトウェア群の不具合など、障害の要因は多様です。計画停止と予期せぬ停止、いずれが発生してもサービスを止めない、あるいはサービスダウンタイムを極小化するITシステムの実現には、しかるべき構築技術が必要です。本稿では、フェールセーフを前提としたITシステム構築技術を概説します。

キーワード

● OMCS　● SI 技術　●高可用性　●フェールセーフ　● VALUMO ウェア

1. まえがき

近年、ブロードバンドや携帯電話の普及により、インターネットへのアクセスが増大してきました。平成17年度の情報通信白書によれば、その利用人口は約8000万人に達しています[1]。一方、総務省が生活者に対して行ったアンケートでは、「安全・安心な生活環境の実現」を求める意見が約7割を占めるに至っています[2]。

このような急速な環境変化に伴い、企業や公共団体などがITシステムによって提供している「使えると便利」だったサービスが、もはや「使えないと困る」サービスへと変化しています。

たとえば、銀行の個人向けサービスは、コンビニATMの設置や店舗営業時間の拡大などにより、時間の利便性や場所の利便性が、近年、急激に展開浸透し、多くの方々に利用されています。また、携帯電話は、音声のみならず、電子メールやWebサイト閲覧、モバイルECなど、その利用形態の広がりから多くの方々に利用されています。

このような多くの方々が利用しているサービスを司るITシステムには、利用者にとって「使いたい時に、使いたいサービスが、期待したレベルで使える」こと、つまり、「ディペンダブル」であることが求められます。

本稿では、ディペンダブルITシステムの構築技術についてご紹介します。なお、ディペンダブルネットワーク構築技術については、Vol.58,No.5に掲載の論文[3]をご参照願います。

2. ITシステムの課題とOMCS

従来、ディペンダブルITシステムの主役は大型汎用機(メインフレーム)やフォールトトレラントシステムでしたが、昨今、オープン製品の採用が増加してきました。外部仕様が公開され、共通化されているオープン製品は、高い価格性能比や先進技術の享受が得られやすい反面、多様な製品群と、そのバージョン間で多数の組合せが存在し、想定外の不具合(落とし穴)が顕著化することも少なくありませんでした。

NECでは、1995年からオープン製品による大規模基幹システム(Open Mission Critical Systems & Solution : OMCS)の構築を開始し現在に至っています。その過程において、オープン製品(best-of-breedプロダクトとも呼ばれる)を用いて、以下のようなSI技術を確立してきました。

- 激甚対策としてのマルチセンター管理、24時間365日無停止無人センター運用を実現した大規模超並列バッチシステム
- 業務無停止でのシステム拡張を可能にした無停止オンラインシステム
- Hubによる多層アーキテクチャと高速サーバ切替技術によるフルオープン勘定系システム
- マルチプラットフォームによる相互運用、リアルタイム経営を支える日次決算システム
- 秒間数万件を処理する超並列スレッド制御ミドルウェア、

数百台の機器の一元管理を可能にしたモバイルインターネットゲートウェイシステム

NECでは、以上のようなシステム構築を通して蓄積してきたSI技術を整理整頓・体系化し、OMCS-SI技術として確立しました。OMCS-SI技術を支える三本柱は、プロジェクト管理技術、アプリケーション開発技術、プラットフォーム構築技術です[4]。

3. ディペンダブルITシステム設計における視点と観点

ITシステムの上流設計では、要件定義が重要とされ、しばしば、機能要件と非機能要件の2つに大別されますが、ともすればサービス内容に直結する機能要件に観点が偏る傾向が見られる場合があります。OMCSプラットフォーム構築技術では、非機能要件をシステム要件として扱い、以下の6つの特性(MC性*)で分類し、それぞれに対して設計評価指標を定義し、整理しています。

1. 高可用性 → サービスを止めない
2. 高性能性 → サービスレベル保証
3. 高運用性 → サービス監視
4. 高連携性 → サービスとプロトコルの直交性
5. 高機密性 → サービスセキュリティ
6. 高拡張性 → 最適投資による段階的拡張性

使いたいときに、使いたいサービスが、期待したレベルで使えるITシステム、すなわちディペンダブルITシステムが具備すべき最も大事な特性の1つが“高可用性”です。これは、あらかじめ決められたサービス提供時間帯でサービスを止めない、あるいは、サービス停止時間を極小化するための機構を作り込むことにより具現化されます。

ITシステムとして「サービスを止めない」ためには、フェールセーフの考え方に基づく障害対策が重要になります。『障害は必ず起きる』という前提に基づき、システム障害の原因を類型化し、システムとして安全側に倒れるような対策を整理します。これら障害原因への対策としては、事前に実施しておくべきもの(事前対策)と、システム障害が起きた場合の処置(事後対策)に大別されます(**表**)。

こうした障害原因の類型化に基づき、技術的な対応策が整理できます。ソフトウェアバグによるシステム障害の場合には、不正アドレス参照によるプロセスダウンやデッドロックによるプロセスストールなどがあります。プロセスダウンの場合には、プロセス監視機構やサービス監視機構により障害発生を早期検出し、プロセス再起動や対象業務閉塞などを自動実行することで障害局所化を実現します。

表　システム障害原因の類型化

項目	システム障害要因	事前対策	事後対策
予期せぬ停止	ハードウェア経年劣化 ハードウェア障害	故障部位に対する予兆検知、ヘルスチェック、二重化・冗長化によるサービス継続	業務閉鎖や縮退で障害部位を切り離し、システムとしてサービス継続
	ソフトウェアバグ	コードレビュー、事前評価	同上
	実装設計ミス	アーキテクチャ／方式設計レビュー	同上
	オペレーションミス	運用手順徹底、訓練、リハーサルでの熟練度向上	早急に緊急対策室を開放し、対応策をトップダウンで検討・実施し、速やかにサービス再開
	スパイクトラフィック、攻撃	なし	トラフィック制御、網規制、最大セッション数・スレッド数で抑制し、全ダウン回避
	激甚災害（天災・テロ）	なし	バックアップセンターに切り替えてサービス継続
計画停止	ハードウェア定期交換	影響範囲・影響度合いなどステークホルダへの事前の周知徹底で混乱防止。 作業対象に応じて、サービス無停止、サービス一時停止、サービス全停止を分類整理し、手順を標準化し、保守時間短縮に努める	
	ハードウェア緊急交換		
	新サービス追加		
	緊急パッチ適用		
	設備系改修		

4. ディペンダブルITシステムの実装技術事例

OMCSプラットフォーム構築技術における高可用性に対する基本的な設計ポリシーは、「疑わしきは切り替えろ**」です。これは、想定される故障部位を監視し、このまま処理が進むとデータ一貫性が崩れるとか、サービス部分停止や全停止など取り返しのつかない状態になると判断したら、即座に対象部位を切り離して縮退するか、待機系に切り替える動作に遷移し、システム全体を安全側に倒すようなフェールセーフの考えに基づいています。

この考え方に対する実現例を、オンライントランザクション処理システム(OLTPシステム)を用いて概説します。

オープンシステムを用いた一般的なOLTPシステムの構造と動作は**図1**のようになります。たとえば、銀行のATMのようなクライアントシステムから送信された要求電文は、アプリケーションノードによって受信され、トランザクションモニタで制御される業務アプリケーション(業務A～業務X)によって処理されます。業務アプリケーションは必要に応じてデータベース

* ミッションクリティカル性：コンピュータシステムが本来提供すべき機能以外に具備しなければならない特性の総称

** 正確には、現用系から待機系への切り替えと、全現用系における切捨て縮退を包含する

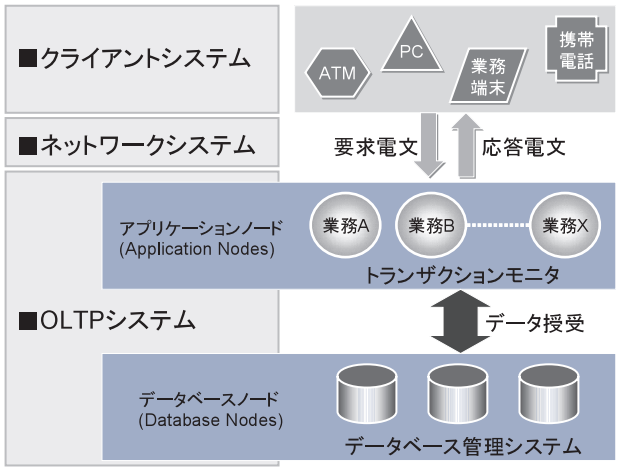

図1　OLTPシステムの構造と動作

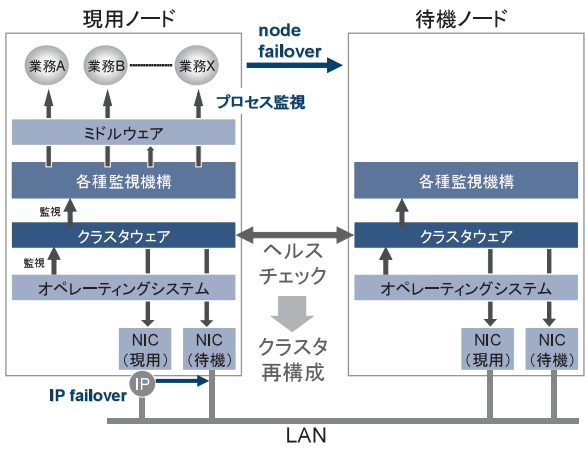

図2　クラスタ構造

ノードとデータ授受を行い、たとえば残高照会のような応答電文をクライアントシステムに返却します。

このとき、データベースノードがなんらかの要因で、業務アプリケーションからの要求が受け付けられなくなった場合、OLTPシステムとしてはサービスが停止し、システム障害となります。このような障害を引き起こす要因には、データベースソフトのダウン、OSストール、OSダウン(PANIC)、CPU故障、メモリ不正、ディスク不良、LAN障害など様々な不具合が考えられます。発生しうる不具合それぞれに対して、予兆の検出機構、発生の検出機構、検出をしかるべき人やシステムに通知する機構、不具合が発生した部位に対する処置、不具合解消後の処置などを設計し実装することが必要です。たとえば、重要なプロセスやタスクに対しては、プロセス監視機構による死活監視やストール監視、プロセスダウン時の自動再起動などが必要です。さらに高度な機能として、一定期間の間(例:5分間)に、再起動を数回(例:3回)実行してもプロセスが再開しない場合は、そのノード自体を強制シャットダウンし、スタンバイノードに切り替える機能があります。

ハードウェアに対しても同様の監視を行います。たとえば、LANカード(ネットワークインタフェースカード;NIC)は、ネットワーク機器とともに現用・待機で二重化し、待機側から現用側にヘルスチェックのパケットを定期的に送信することにより監視します。ネットワークに不具合が発生し、現用側が停止した場合、待機側から送信されたヘルスパケットが途切れます。その場合、現用側のNICを非活性化し、現用側に割り当てていたIPアドレスを、待機側のNICに割り当て、活性化します(図2;IP failover)。通常、IP failoverは、数秒で完了するため、通信相手に対しては、NICなどネットワークの不具合は隠ぺいされます。

複数のノードでクラスタを組むことにより、サービスダウンタイムを短くすることができます。クラスタを構成するノードは、相互にヘルスチェックを行うことで不具合の発生を検出します。ノードダウン(PANIC)やOSストール、重要プロセスダウンを契機にヘルスチェックが途切れ、不具合の発生したサーバがクラスタから切り離されます(図2;クラスタ再構成)。通常、クラスタ再構成には数十秒～数分の時間が必要となります。クラスタ再構成後、アプリケーションの再起動やリカバリが行われるため、サービスの再開までには数分～十数分の時間が必要となります(図2;node failover)。したがって、クラスタ再構成が伴う不具合発生時には、クライアントにサービス停止が見えてしまうことがあります。

ディペンダブルITシステムの高可用設計では、このサービスダウンタイムをいかに短くするかが重要となります。たとえば、OLTPシステムを構成するデータベースノードでOSダウン(PANIC)が発生してからサービスが再開されるまでのサービスダウンタイムは、通常、数分～10分となりますが、システムによってはサービスダウンタイムの許容値が60秒以内である場合があります。そのときは現用・待機型のDBノード切り換え技術では実現が難しく、ホットスタンバイ的な技術が必要となります。

また、複数のサーバから1つのデータベースを共有できる製

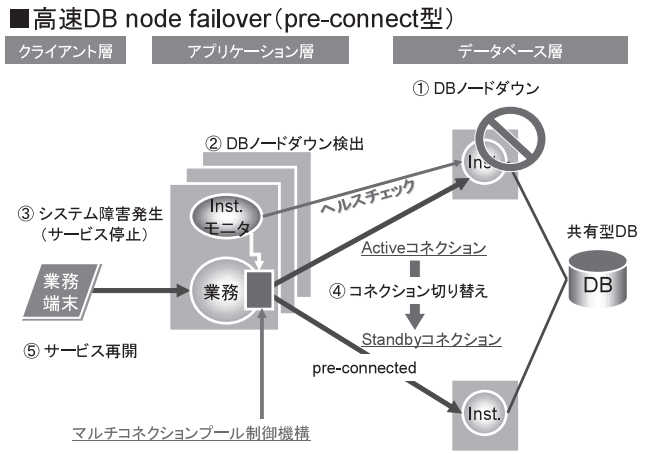

図3　共有型DBを用いた高速DB切り替え技術

品を採用しアプリケーション側から擬似的に現用・待機として使う技術があります(図3)。

ただし、この実装には、

・何をもって現用サーバーダウンと判断するか
・Standbyコネクションに切り替える契機の取り方
・アプリケーションサーバがダウンしたとき、処理中のトランザクションがロックしていたレコードをどうやって解放するか(残存ロックの強制刈り取り)

など考慮すべき設計ポイントがいくつか存在します。

NECでは2000年に、上記高速DB切り替え技術を完成させ、いくつかの実システムへの適用を行ってきました。製品によっては、データベース層内での障害検出機構が提供されている場合もありますが、ここでは「アプリケーションから見てデータベースが利用可能か」という観点がポイントです。よりマクロなレベルでは、擬似的な業務プログラムからデータベース層まで監視も可能ですが、その場合、検出可能なサービスダウンのスコープが広がる一方、サービスダウンタイムも長くなります。サービスダウンの極小化や局所化には、構成要素、サブシステム、システム全体といった階層ごとの障害の検出と、それぞれの検出機構の連携が必要になります。

このような高度なディペンダビリティを持つシステムは、best-of-breedプロダクトだけでは実現が難しく、その不足分を補うミドルウェア(補完ミドルと呼ぶ)が必要となります。NECでは、これまでのシステム構築を通して現場で作り込んできた補完機能のなかで、汎用性の高い機能を、順次VALUMOウェアとして製品化してきました。

これにより、高い信頼性を具備した大規模なITシステムを、安全・確実・スピーディに構築してきました。

特に、高可用クラスタ関連のVALUMOウェアは、第2章で挙げたSI事例をはじめとする、高い信頼性を要求されるお客様のITシステムで採用されてきました。

5. むすび

今後のユビキタス社会では、さらに多彩なビジネスモデルが生まれ、IT-ネットワークシステムとして具現化されます。その進化に伴い、IT技術とネットワーク技術の融合が進み、今以上に重要インフラとしてのディペンダビリティを求められることになると認識しています。

本稿では、特に、ITシステムのディペンダビリティの考え方と構築技術を中心に紹介しました。NECでは、今後もOMCS、NGN、IT-NW統合ソリューションを軸にSI技術を研鑽し、ユビキタス社会を支えるサービス事業を推進していく所存です。

参考文献

1) 総務省;情報通信白書;
http://www.johotsusintokei.soumu.go.jp/whitepaper/ja/cover/index.htm
2) 総務省;「ユビキタスネット社会の実現に向けた政策懇談会」最終報告書;
http://www.soumu.go.jp/s-news/2004/041217_7.html
3) 阿留多伎ほか ;「ディペンダブルネットワーク技術による情報通信インフラ構築」, NEC技報, Vol.58, No.5, pp.79-85, 2005-9.
http://www.nec.co.jp/techrep/ja/journal/g05/n05/g050528.html
4) 富山ほか ;「オープンミッションクリティカルシステム構築技術」, NEC技報, Vol.56, No.7, pp.47-50, 2003-8.
http://www.nec.co.jp/techrep/ja/journal/g03/n07/g030712.html

執筆者プロフィール

東 健二
官庁・公共・金融・通信ソリューション企画本部
統括マネージャー

上窪 真一
通信・メディアソリューション事業本部
シニアマネージャー

安場 洋輝
官庁・公共・金融・通信ソリューション企画本部

●本論文に関する詳細は下記をご覧ください。
関連URL: http://www.sw.nec.co.jp/solution/omcs/