社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2006－EVA－17 (8)
2006／6／20

災害対策システムの構築技術と要件検討手順

高山 晴禎*

The construction technology and the requirement review procedure for a Disaster Recovery system.

Harusada Takayama*

今日では基幹業務システムの 24 時間 365 日連続稼動の要求は高まる一方となっており，大震災のような大規模広域災害発生時においても短時間で回復・稼動出来ることが求められている．長距離間でストレージの同期を実現する代表的な仕組みとして PPRC (Peer-to-Peer Remote Copy)グローバルミラーが挙げられる．この方式では予備局はスレーブとなり，密接な結合方式となる．一方各拠点のコンポーネントが独立して動くと同時に全体としては協調するフェデレーション（連邦）という考え方を実現する方法がある．ミドルウェア製品やアプリケーションを活用してデータ整合性を確保していこうというアプローチである．ここで重要なのはサービス要件達成のために最適なシステム形態をどのように判断して選択するかという点である．本稿では，密結合方式と疎結合（フェデレーション）方式の対比を通して，高可用性実現の要件検討手順を明確にする．

The demand for continuous operation of the mission-critical system 24 hours a day, every day is now steadily growing. It is required to recover and resume its operation in a small amount of time when a large-scale, wide-area natural disaster like a great earthquake has occurred. Peer-to-Peer Remote Copy (PPRC) global mirroring can be cited as a typical mechanism for realizing the synchronization of storage as a remote distance. In this case, a tightly linked backup station is used as a slave. On the other hand, there is a method for creating a federation, in which its component functions located at different sites perform independently while they can also collaborate as a whole. This consistency of data is ensured by utilizing middleware products and applications. What is important is how we should evaluate and select the optimum system structure for archiving the service requirements. Through a comparison between a tightly linked system and a loosely linked (federation) system, this paper attempts to clarify the requirement examination procedures for the realization of high availability of the mission-critical system under construction.

Key Words & Phrases ：災害対策，高可用性，データ保全性，PPRC，ミッションクリティカル
Disaster recovery, High Availability, Data Integrity, PPRC , Mission Critical

１．はじめに

基幹業務システムの24時間365日連続稼動を実現する方法としてはクラスタリング技術による 1 拠点内での待機系への切り替えの仕組みが多く採用されているが，拠点全館停電などの計画停止要件が発生した場合や，大震災などの大規模広域災害が発生した場合には対応できないため，本番局・予備局 2 拠点体制[1]による高可用性実現方式が必須となる．

2 拠点間のデータの整合性をどのような方式で確保するかについては大きく 2 つの方式に分けられる．1 つ目はストレージのインフラ技術によるリモートコピー方式で，ここ数年で SAN ストレージ技術，IP 網やダークファイバー技術の向上が見られ，ストレージ技術で整合性を確保するという方式が実現出来るようになった[4]．長距離間でストレージの同期を実現する仕組みとしては PPRC (Peer-to-Peer Remote Copy)グローバルミラーが代表的な方式である．この方式では本番局はマスター，予備局はスレーブとなる．2 つ目はミドルウェア製品やアプリケーションのレプリケーション機能を使って整合性を確保するというフェデレーション方式で，最新の製品で機能向上が図られている．フェデレーション方式は各拠点のコンポーネントが独立して稼動するという特徴を持つので，ストレージ技術を利用する方式ほどは本番局に依存しないデザインを採用することが可能である．ストレージ技術を使う方式が密結合であるとすれば，フェデレーション方式は疎結合と言える．

ここで重要なことは先ず製品ありきで設計するのではなく各方式をどのような要件の際に適切に用いるかということを最初に検討することである．本稿では密結合方式と疎結合方式との対比を通して，高可用性実現の要件検討手順を明確にする．以下第 2 章では災害対策に求められる災害対策レベルと製品の実装レベルを紹介し，システム構想段階での両方式の選択基準を述べている．第 3 章では高可用性実現のための要件検討手順を述べる．

*e07593@jp.ibm.com，日本アイ・ビー・エム株式会社 通信・メディア・公益事業
Communications Sector, IBM Japan, LTD

2．システム構想段階での検討項目

2.1 災害対策レベルの検討

災害対策はコストを掛ければその分高度な技術を利用することが出来る．また災害回復までの平均リカバリー時間が短いソリューションを構築出来る．米国 Share は災害対策レベルと現状での災害時の対応を表 1 のようにまとめて位置づけしている．災害対策レベルとは災害対策の評価指標であり，対策無しのレベル 0 からデータロス無しに遠隔地で業務を引き継げるレベル 6 までの 7 段階に分けて規定されている．災害対策システムの構想の段階で最初に考えるべき事項として下記が挙げられる．

① 災害対策レベルをどこまで実現するか
② RTO(Recovery Time Objective どの程度の時間で切替出来ることを目指すか)
③ RPO(Recovery Point Objective どの程度前までのデータを確保するか)
④ 結合方式（疎結合にするか密結合か）

これらの検討を実施している段階では，正確な数値を出すというよりも目標値として大まかな RTO, RPO の値を設定し，目標達成のためにどの程度の対策レベルを実装していけば良いのかを決めていく．具体的製品名はこの後の段階で検討していくほうが効果的である．

表 1．災害対策レベル（Share にて制定）[4]

対策レベル	現在のシステム構成	災害時対応
0 遠隔地ﾃﾞｰﾀ無	災害対策用の遠隔地保管データは存在しない	システム用施設の対災害性の強化
1 PTAM (ﾋﾟｯｸｱｯﾌﾟﾄﾗｯｸ方式)	遠隔地にデータを保管．一般的にはテープにデータを入れて陸送．	復旧用システムの手配 テープよりシステムを回復
2 PTAM +復旧用ｻｲﾄ	PTAM に加え，復旧用のサイトを保持している	復旧用システムを利用してテープよりシステムを回復
3 日次電送方式	2 に加え，一部重要なデータは日次で電送．その他 PTAM で陸送	復旧用システムを利用して電送データとテープよりシステムを回復
4 予備局常時活動方式	分割された 2 つのサイトで業務を実施．PTAM と非同期伝送でデータを互いに保持	復旧サイトで電送データとテープよりシステムを回復
5 2 フェーズコミット方式	本番システムと復旧用システムを持つ．処理は復旧用システムへのデータの書き込みを以ってコミットする．システム間の二重コピー	復旧用システムを利用して即時に再立ち上げ
6 データロス無し	遠隔地システムとローカルシステムが単一システム・イメージで稼動．互いにデータを二重コピーで保持し，かつデータの共用性も維持	システム全体で自動的に処理を行う．システムをスイッチし適用業務は停止しない

2.2 製品の対策レベルと結合方式

密結合の例として，PPRC グローバルミラーと DBMS の Log Shipping が挙げられる．PPRC グローバルミラーは頻繁に差分フラッシュコピーを取得出来る機能があるので，対策レベル 4 ではあるがデータロスは少ない実装レベルにある．Log Shipping は容易に遠隔地への災害復旧が出来るようになっており，対策レベルは 5（同期モードの場合）となる．

疎結合方式の例としては DB Replication やアプリ転送等が挙げられる．アプリケーションでの転送は最小単位としてテーブルやデータの一部のみを扱うことが出来る点が特徴となり，より柔軟な処理を実現することが出来る．疎結合方式は予備運用局で DB が立ち上がっており，予備運用を実施しながら，本番業務等を同時並行的に実行することが可能な点が特徴である．

密結合方式には，次のような特徴がある．
①本番局と予備局のストレージが主従関係
②従側のストレージを従側 OS からアクセス不可能
③対象ストレージの更新内容全てが反映

一方，疎結合方式には次のような特徴がある．
①本番局と予備局は一定間隔で同期を取り，主従関係は明確には設定しないのが原則
②予備局のデータに予備局 OS からアクセス
③本番局更新データの全てを反映させない

2.3 災害対策レベルと密結合・疎結合の選択

製品の選択に入る前の構想段階で目指している災害対策システムでは何を目標とするのかについて次の 2 点から検討することが効果的である．ハード，ソフトの製品ありきで最初に検討を進めてしまうとバランスが崩れ，特定の機能ばかりが優れているが全体としてのレベル低下を招くことが考えられるからである．

① どこまでの対策レベルを実現するか
② 密結合にするか，疎結合にするか

①については災害対策レベルが高いほど，短時間で最新のデータに戻すことが可能で，対策レベル 6 を目指すシステムを作れば瞬時に遠隔地で業務を再開させることも可能となる．ただし対策レベルが高いほどシステム構築にかかわるコストは高額になっていくので，概算でトータルコストを考えながら目標とする対策レベルを定義していく。②については明確に決定するのは第 3 章で述べる要件の検討を進めていくことで結論を出していくことになるが，この段階で

は目標として密結合・疎結合どちらの方向を中心に検討していくかを定義していくことが今後の設計を進めやすくすることになる．3 章以降で具体的製品選択をしていく前段階として，製品の持つ基本機能を理解していくため結合方式をＸ軸に，対策レベルをＹ軸に製品をプロットし，要件検討の際に利用していくと良い．複数の製品を組み合わせる可能性がある場合には 2 つ以上の対策レベルの離れた製品同士を組み合わせると全体としてのバランスが崩れてしまうので注意が必要である．

2.4 災害対策システムの要件検討手順

図 1 は災害対策システム構築の際の要件検討手順を示したものである．システム構想の後の要件定義フェーズでは通常は本番局，予備局間の距離とインフラ（H/W,S/W,N/W）を検討した後でシナリオ・ベースを検討していくという手順を取るが，シナリオの範囲がはじめからある程度明確になっている事例の場合はインフラ要件よりも先にシステム設計要件を検討することによって切替シナリオを明確に設計し，シナリオに最適なインフラ要件を検討していくという方式も有効である．

密結合・疎結合のどちらの方式を採用するべきかについての判断を行う場合密結合方式を先ず優先して検討すると要件を明確にしやすい．ネットワーク速度が業務量にてらしあわせて余裕がある場合や切替シナリオを一度作成したら追加変更が少ないケースでは密結合方式が，シナリオの追加・変更が多く発生する場合やデータの入れ替わりが激しい業務を対象とする場合は疎結合方式の採用を優先して考えていくほうが望ましい．

表 2 は密結合方式の代表である PPRC グローバルミラーと，疎結合方式の代表であるミドル AP

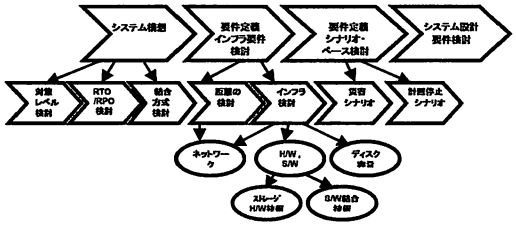

図 1..災害対策システムの要件検討手順

表 2. 局間転送方式の比較

比較対象	密結合方式 (PPRC グローバルミラー)	疎結合方式 (ﾐﾄﾞﾙ APL ﾘﾓｰﾄｺﾋﾟｰ)
災害対策レベル	4 (但し同期ﾓｰﾄﾞ実装により 6 も可能)	4 (但し同期ﾓｰﾄﾞ実装により 6 も可能)
DB の局間整合性	大	中
災害時の回復方法	ディスク活動化+クラッシュリカバリー	予備局業務稼動
コスト	大(Network 大域確保と DWDM 設備)	小
データロスの可能性（災害時）	3 ～5 秒前以内の変更と同期点	非同期による転送処理分
運用時の考慮点	ﾃﾞｨｽｸ中身が大幅に変更されると反映に時間が掛かる	ｱﾌﾟﾘ仕様に変更があるたびにﾃｽﾄが必要
転送不具合発生時の処理	ｸﾗｯｼｭﾘｶﾊﾞﾘｰ	ケースに応じて各種ﾂｰﾙを作成して提供

リモートコピー方式の機能比較を行ったものである．各比較事項の詳細については第 3 章以降で述べていくものとする．

3．要件定義・設計段階での検討項目

3.1 インフラ要件の検討

高可用性要件を検討していくにあたり，先ずは以下に述べるインフラ要件から総合的に検討を進めていく必要がある．

① 本番局，予備局間の距離
② ネットワーク速度
③ H/W が提供する機能と災害対策レベル
④ S/W が提供する機能と災害対策レベル
⑤ ディスク容量

①は同期モードで予備局転送出来るか否かに影響を与える．同期モードで転送出来れば災害対策レベルは 5 または 6 を実現可能となるが，距離が短いと大規模広域災害に対応出来なくなる（両拠点壊滅）という短所もあるので，どちらの要件を重要視するかを決定しておく必要がある．

②はコストおよび距離との兼ね合いとなるが，可能であれば専用に安定した速度，かつ高速のネットワークを利用出来るようにすることが望ましい．コスト上難しい場合は密結合方式よりも疎結合方式を優先して検討を進めていく必要がある．その理由については 3.4 で述べていく．

③④については 2.2 および 2.3 節で述べている対策レベルを製品毎に詳細検討し，実現しようとしているシステムでどの対策レベルを実装していくかを決定していく．この段階においても，

具体的製品を限定して検討するのではなく，密結合方式，疎結合方式という 2 方式を意識しながら検討を進めていくほうが，3.4 のシステム要件を検討した結果別の仕組みに立ち戻る際に戻りやすく，再検討しやすくなる．

⑤のディスク容量については密結合の代表的な方式である PPRC グローバルミラー方式を採用する場合等には実容量の 3 倍ないし 4 倍以上のディスクが本番局，予備局トータルで必要になるので，H/W 選定にあたり考慮が必要となる．

構築予定の災害対策システムが求めている対策レベルを検討し，①から⑤の要件を検討して全体に整合性を持つシステム設計を行うことが高可用性システム実現に向けての要件定義段階第 1 ステップとして必要である．

3.2 シナリオ・ベースの検討

連続稼動システム基盤を構築する際に考慮すべき要件として「計画停止要件」と「災害停止要件」を導入した．

「計画停止要件」とは，24 時間 365 日連続稼動を実現するために障害・災害以外事象を原因としてデータロスが発生することなく予備局に切替を行うことが特徴であり，以下のような例が挙げられる．

① 拠点全館停電により 1 日で本番局に復帰
② 停電を伴わない本番局保守作業
③ ハード保守による予備局 1 日停止
④ 定期的な予備局への遷移（運用目的）

「災害停止要件」とは，H/W,S/W,N/W の障害発生を原因としたり，災害が発生した時にある程度のデータロスを伴いながら予備局に切替を行うことが特徴であり，以下のような例がある．

① クラスター機能でも対応不可能な障害
② 大震災のような大規模広域災害
③ ネットワークの関連する全面障害

上記 2 要件は似ているが細分化していくと相反する要素を持っているケースがある．例えば災害が発生した時には災害発生局の復旧はある程度時間を掛けて行うことが可能であるが，計画停止の場合数時間から 1 日のうちには停止した局に復帰できる要件があるケースが一般的である．そのため，本番局と予備局の方向を短時間で逆向きに設定する必要が生じ，且つデータの差分のみを送信出来るようにする機能が求められる．図 2 は本番局と予備局を設けてその間でどのような整合性確保の仕組みを構築するかを概念的に現したものである．

ここで重要なのはシナリオ・ベースでデザインを行っていくことで，全てのシナリオでディスク配置に関しては合理性を持って検討されていく必要があるとの知見を得た．特に災害切替の場合発生時期が予想不可能のため誰でも手順書通りに出来ることが重要で，そのためシナリオは多くの災害ケースに対応していなければならない．代表的なシナリオとして下記が挙げられる．これをさらに細分化していくことで要件が明確になり，具体的実装が可能となる．

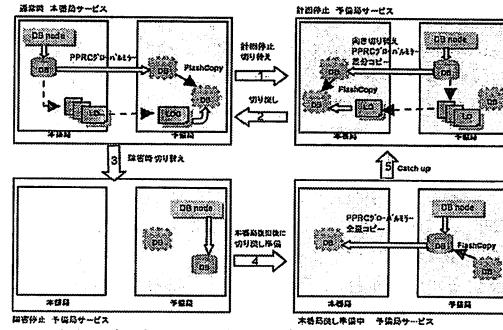

図 2.本番局と予備局データ同期の仕組

(1) 本番局から予備局への計画切替(1 の矢印)
(2) 予備局から本番局への計画切戻し(2 の矢印)
(3) 本番局から予備局への災害切替(3 の矢印)
(4) 予備局から本番局への災害切替
(5) 本番局本番運用時の予備局メインテナンス
(矢印では 1 の差分再送に該当する)
(6) 予備局本番運用時の本番局メインテナンス
(矢印では 2 の差分再送に該当する)

図 2 の仕組みで計画停止要件(1,2 の矢印)に加え災害発生時の要件(3,4,5 の矢印)の両方を満たすことが可能となる．図 3 は計画停止の概念図を詳細化した図となる．本番局内部でクラスター片寄せ運用では対応不可能な計画停止要件が発

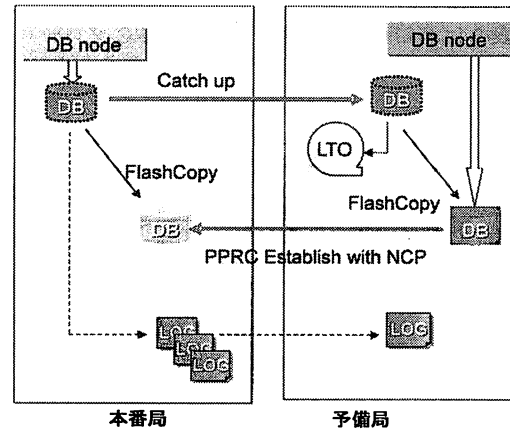

図 3．計画停止の概念詳細例

生した場合にシナリオが実行される．図 3 のケースでは論理的な DB は本番局に 2 つ，予備局に 2 つ存在している．

3.3 システム設計要件の検討

高可用性要件を検討していくにあたり，シナリオ・ベース要件を検討した次の段階ではシステム設計要件について下記の観点から総合的に検討を進めていく必要がある．

① 局間切替シナリオの範囲
② 局間データ整合性の時間精度
③ 全停止までの最長時間
④ 縮退運用時の局間切替への考慮
⑤ 予備運用局の可用性への考慮
⑥ テープバックアップのタイミング
⑦ テープバックアップの取得拠点
⑧ 運用担当者のスキルレベル

①は 3.2 で詳細を論じているが，災害切替のみをシナリオ対象とすれば良いのか，計画停止を加えるのかという検討を業務要件から検討していく．②は本番局のデータが予備局に反映されるのにどの程度前まで保証されているべきかを検討する．③はフラッシュコピーの物理コピー（OS から見ると瞬時にコピーされるが実データのコピーには時間を要する）が終了する時間まで全館停電のような法廷点検は実施できないため，運用上何時間まで稼動していることが許されるのか等の要件を決めておく．

④は災害切替をしなければならない事態に陥る前に既にクラスタリング機能で局内内部切替が発生していることを前提に，局内片肺状態での局間災害切替シナリオを想定するかしないかを検討する．二重災害に思えるが実際のケースでは発生する可能性が高いシナリオなので，採用可否を検討していくことを提案する．

⑤は予備運用を行っている局内部でノード障害が発生した場合に，HACMP を稼動させて切替を行わせるべきかの検討である．予備運用局の可用性を高めることで本番運用局の災害発生時のデータ損失を最小限に抑えることが可能である．

⑥⑦については，テープを本番局・予備局のどちらで取得するべきかと取得タイミングを検討していく．

⑧に関しては，3.2 で述べたとおり特に災害切替シナリオを実行する際には 1 つのシェルスクリプト実行により一連の切替が全て完了するようなデザインが望ましいと考える．

以上のシステム設計要件を検討していくと，3.1 で述べたインフラ要件に立ち戻る状況が発生する場合がある．システム設計要件とインフラ要件は密接に関連しており，システム設計要件を追求していくとより多くのディスク容量が必要になるなど，インフラ要件が変わることがある．特に密結合方式を採用した場合は 1 つの要件が他の要件に与える影響が強くなる傾向がある．

3.4 密結合方式による設計例と考慮点

図 4 は PPRC を利用した密結合方式によるシステム設計の例である．図中の Controller オブジェクトが他のミドルウェアを制御し，一括運用を行っている．どのシナリオを適用対象とするかについては図 4 のような設計を大きなシナリオを分割して決定していく．密結合方式設計の際には各サブシステムが密接に関連し，順序性も要求されるため，図 4 のようなオブジェクト単位での設計を行うことが要件を明確にしていく上で重要である．Controller オブジェクトは最終的には OS 上で稼動するシェルスクリプトという形で実体化し，整合性を持って稼動していくようになる．

ネットワークに負荷を与える再編成の影響は密結合方式を利用する方式では考慮することを提案する．再編成する運用方針がある場合は注意が必要で，その場合再編成処理による更新量とネットワーク能力を計算し，どの程度の遅延が発生するかを検証していく必要がある．

3.5 疎結合方式における設計例と考慮点

ミドルウェアで提供される Replication の機能やアプリケーションの 2 重書きの機能を使えば密結合方式に比較して双方向転送は比較的容易に実現可能であり，システム要件の変更についても容易に対応できる．密結合方式はクラスタ

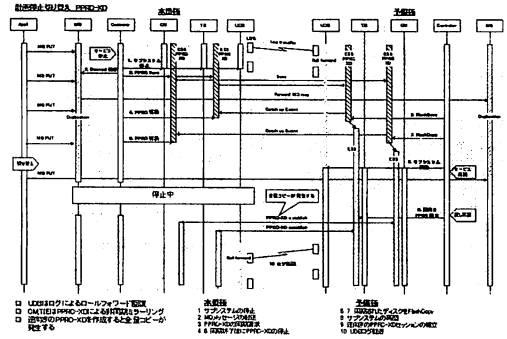

図 4.PPRC のシステム設計例

一機能の起動・停止の順序性とも密接に関係を持つので，シナリオに修正があった場合のテストの負荷については高くなる傾向にある．そのため「疎結合」で本番局と予備局を結びつけるフェデレーション方式も有効な災害対策の 1 手法として注目されるようになってきている．Replication など DB から提供される機能によって，フェデレーション方式も精度の高い仕組みが構築出来るようになってきた．

フェデレーション方式の旧来からの代表的な手法としては，アプリケーションによる 2 重書きが存在する．必要なデータのみを抽出するためネットワークへの負担が小さく，データの論理的な整合性を確保しやすくなるというメリットがある反面，アプリケーションや DB の仕様追加・変更がある度に仕様の変更やテストが必要となり，また 2 重書きアプリケーションの一部機能不全があると，一部は更新されたが一部更新できていないデータが発生するなど，障害発生時の問題判別に時間と手間が掛かるというデメリットもある．

図 5 はミドルウェア製品を利用したリモートコピー方式の設計例である．このアプリケーション転送方式では，以下の点を考慮することが重要である．

① 日次一括削除などのバッチ更新が大量に行われると転送キューのキャパがフルになるためそのような処理は各局にて行う．
② 転送用ストレージ（ファイルシステム格納用）と転送キューは，各局が停止している想定時間の間蓄積されても問題にならない量を確保する
③ 転送キューは転送効率を考慮して複数準備する
④ 複数のキューでの更新順序に対する整合性確保を行う

これらの考慮点はアプリケーションによる 2 重書きや Replication を活用する際にも同様に考慮されるべき課題となる．Replication を利用する

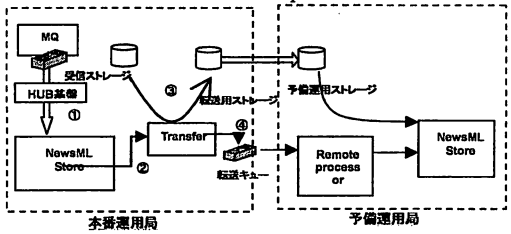

図 5. ミドルウェア製品リモートコピー方式

方式では，(3)(4)の課題に対する整合性確保のための工夫として，図 6 が挙げられる．転送キューを複数にして予備局側の処理を並行化し，処理速度を高めている．また，同じ Public ID の処理順序を保持するようにデザインする（同じ Public ID のメッセージは同じキューに積むようにする）ことで，順序性を維持しながらパフォーマンス向上を図ることが可能である．

アプリケーションの機能不全等によりデータの整合性の問題が発生する課題への対策としては，ソース表とターゲット表を比較するツールや，表の修正やコピーを行うツールを提供することで解決可能である．

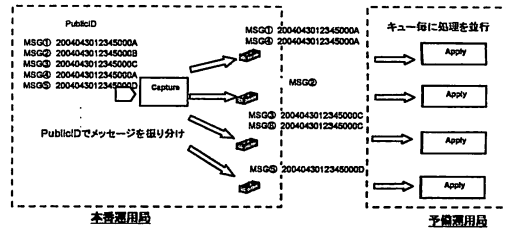

図 6. Replication 利用時のメッセージ振分

4．おわりに

本論文では，密結合方式と疎結合方式の技術について実際の設計例と実装例を提示しながら要件検討手順について論じてきた．メリット・デメリットを理解し，設計においては製品ありきではなく方式検討が重要である．本システムの実装技術は多くのミッションクリティカルな業務に展開可能な技術であると考える．

参考文献

[1] 加藤礼基，遠隔コピー機能による災害対策システム構築に関する考察,情報処理学会研究報告,2003 年 5 月 vol.2003 no.061
[2] 石田修/瀬戸康一郎，10 ギガビット Ethernet 教科書，IDG ジャパン，ISBN4-87280-460-0，2002 年 4 月
[3] Ian Foster, Carl Kesselman, The Grid:Blueprint for a New Computing Infrastructure, Morgan Kaufmann, 1998
[4] IBM Redbook Total Storage Disaster Recovery,2004 http://www.redbooks.ibm.com/abstracts/sg246547.html
[5] IBM Japan, http://www.ibm.com/jp/, 2004.8.28
[6] 情報通信総合研究所編，情報通信アウトルック 2002,NTT 出版,2002
[7] Thomas H. Davenport, ミッション・クリティカル，ダイヤモンド社,ISBN4-478-37345,2000