ハイブリッドクラウドにおける データベースマイグレーションの性能に関する考察

理学専攻　情報科学コース　西出 彩花　(指導教員：小口 正人)

1　はじめに

近年，コンピュータシステムにおける情報量が爆発的に増加している．その処理プラットフォームとして，ハイブリッドクラウドの利用が注目を集めている．ハイブリッドクラウドを利用する際には，パブリッククラウドとプライベートクラウド間で，データベースの冗長的な同期が必要である．また，日本は火山やプレートに囲まれており，地震などの自然災害の影響を受けることが多い．2011 年の東日本大震災で多くのデータが失われたことなどにより，災害発生時には迅速にデータを保護する必要性が強く認識された．そこで本研究では，データベースの冗長的な遠隔バックアップと，災害時の外部情報をトリガとするマイグレーションによって継続的なデータアクセスを実現するシステムを提案する．

2　関連研究

2.1　Pangea

ここで本研究で用いる Pangea[2] について説明する．Pangea は，NTT 研究所で開発されている，LAN 環境を前提としたデータベース同期ミドルウェアである．コンシステンシを保ちながら，複数台のサーバでデータベースのクエリ処理を並列実行する事により，性能向上を実現した．これら複数台のサーバは，同一のデータベースイメージを保持しており，同一 LAN 上に接続されている．サーバの 1 台を Leader，その他は Follower としており，クライアントからこのミドルウェアを介しサーバにアクセスして同期をとる．全てのクエリは照会処理と更新処理に分類され，照会処理はどれか 1 台のサーバで，更新処理は全てのサーバで実行される．更新処理の場合は Leader に対して更新をした後に，Follower に対しても同様に処理を行う．

3　緊急災害時の動作

緊急地震速報のような外部からの緊急災害や負荷変動を予告するシグナルが入って来る際に，それをトリガーとしてシステム構成を切り替えるシステムを考える．

まず，トリガとなる外部情報を取得する．この手法として，災害の情報を Twitter などの外部情報から検知することが挙げられる．例えば，大地震後によるデータセンタ本体の物理的損傷対策の場合，気象庁が発表する緊急地震速報の BOT は，投稿される形式が決まっているため，そのツイートから Twitter Stream Reader を用いて，震源地や震度や地震の規模を取得することが可能である．これを利用して，大地震が到達する前に，仮想マシンを遠隔地にある物理サーバへマイグレートすることで，大地震によるデータ損失の被害を防ぐことが可能だと考えられる．

4　実験

4.1　OpenStack への Pangea の導入

OpenStack[1] を用いて仮想環境上にクラウド基盤を構築し，インスタンスのマイグレーションを行うことにより，提案システムを評価する．クラウド上で Pangea を用いる性能考察を行うために, 構築したクラウド環境内に仮想マシンを立ち上げる．OpenStack とは，クラウドを構成する仮想マシンや物理サーバの運用管理を実行し，それを効率的に行うためのオープンソースのクラウド構築ソフトウェアである．複数のコンポーネントから構成され，これらのコンポーネントが連携することで IaaS のサービスを提供する．

OpenStack においては，コントローラノードから指示を出し，この指示に従ってコンピュートノード上でインスタンスが起動される．本実験では OpenStack のコンピュートノード 4 台のうち，1 台をクライアントサーバ，1 台を Pangea 配置サーバ，2 台この際，仮想マシンのインスタンスとしては，ダウンロードした Ubuntu14.04 のインストールイメージを元に，80G のディスク領域を使用した Linux OS が立ち上がるように設定する．をデータベース配置サーバとする．

データベース配置用のサーバに PostgreSQL サーバをインストールし，クライアントサーバでは TPC-W 用の Tomcat を動かし，Pangea 配置用のサーバ上で Pangea を立ち上げる．今回はこの環境下でデータベース間での転送を行い，スループット値やレスポンス時間を計測することで，Pangea の利用による転送の性能の低下が起こらず Pangea の導入が有用であることを確認する．

4.2　実験結果

前章で紹介した環境下で性能検証を行う．TPC-W における仮想ブラウザ数 EB（クライアントからの負荷量）を変化させた時のスループット値は EB 数 600 の時，78.24 であった．インスタンスを介さず物理サーバ上で直接同じ実験を行った結果である EB 数 1000 の時の 126.78WIPS と比較すると，最大スループット値が 39%低下するなどの大きな性能の低下がみられる

.

ここで，データベースの転送の際に最もボトルネックとなるのは，仮想マシンのデータベースがノード自身の物理データベースへアクセスする際にコストがこ

とである．そのため，よく使うデータベース上のデータはバッファプール上にまとめておいた方がアクセスが速くなると考えられる．

Postgresql はトランザクションログを作成するが，これはデータベースの全ての更新内容をログとして保存し続けるため，データ量が多く，ハードディスクの I/O 負荷を高くする．このため，新たなデータベース用のテーブルスペースとしてシステム上のデータベース部だけを物理サーバに置くことで、デフォルトテーブルスペースにはできるだけデータベースを作らずに，トランザクションログのための I/O 負荷をクラウド上のデータベースに，データベースアクセスのための I/O 負荷は物理サーバ上に配置することで，I/O 負荷の分散が期待される．このイメージを図 1 に示す．

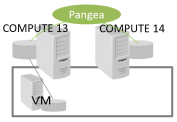

図 1: データベース共有のイメージ

これらより，データベース用のサーバを物理サーバ上におくことが有用であると考えた．これは，データベースを物理サーバ上に置くことで，緊急時以外にクラウド上に重要なデータを置かずに済むため，セキュリティ面を考えても有用なシステムだと考えられる．この時，最大スループット値は EB 数 900 の時に 122.1WIPS となり，システムがすべて物理上に置かれた時と比べても約 4%の性能低下に抑えることができた．本研究で行った 3 種類の環境での実験結果を図 2 で比較した．この結果から，仮想マシンが用いるデータベースを外部の物理サーバ上に置くことで，インスタンスをクラウド上に置く際におこる仮想化によるオーバヘッドを軽くすることができると考えることができる．

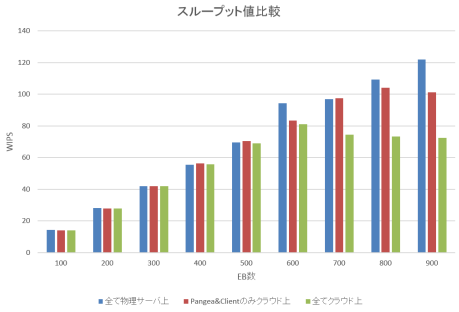

図 2: システム構成によるスループット値の比較

5　提案システム

これらより，OpenStack を用いて構築したクラウド環境上で，データベースの遠隔バックアップを冗長的に動かしながら，緊急災害時にインスタンスのマイグレーションを行うことにより，提案システムを実現した．本提案システムのイメージを図 3 に示す．用意した 2 つのコンピュートノードのうちの 1 つに仮想マシンを構築する．仮想マシンは，システム以外のデータはコンピュートノード上で管理する．2 つのノードは Pangea により常に同期されている．仮想マシンがマイグレーションされる際に Pangea の接続は解除されるが，転送中も Pangea が置かれているサーバ上のデータベースが起動しているため，転送後に差分を抽出すれば，仮想マシン転送中に受けるアクセスについても対応することができる．

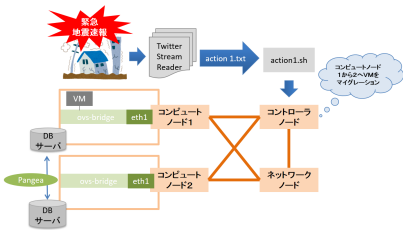

図 3: 提案システムのイメージ図

6　まとめと今後の課題

OpenStack(Icehouse) を用いて構築した環境内での仮想マシンマイグレーションの性能検証を行った．本稿では，データベースサーバ内の全てのデータを同期することを前提としている．しかし，セキュリティの問題などから，実際にハイブリッドクラウドとしてクラウドを利用する際にはパブリッククラウドとプライベートクラウドに保存されるデータは一部異なることが期待される．そのため，要求されたクエリがパブリッククラウドからのアクセスが認められたデータであるかどうかも Pangea 内で区別することにより，テーブルごとのデータ管理を可能としていきたい．これは，Pangea がパケットを元にクエリの種類の判断をしていることから，可能であると判断される．

謝辞

本研究を進めるにあたり，ＮＴＴ研究所 細谷柚子様に数多くの助言を賜りました．深く感謝いたします．

参考文献

[1] OpenStack : http://www.oprnstack.org/

[2] T.Mishima and H.Nakamura : "Pangea:An Eager Database Replication Middleware guaranteeing Snapshot Isolation without Modication of Database Servers", Proc.VLDB2009,pp.1066-1077, August 2009. PVLDB2009.

[3] 原瑠理子，小口正人：「緊急災害情報に基づく OpenFlow を用いたバックアップシステムの実装と評価」 第 8 回データ工学と情報マネジメントに関するフォーラム (DEIM2016)，E7-5，2016 年 3 月．