さくらのクラウド・ストレージに関する報告書

2012 年 6 月 25 日
さくらインターネット株式会社
代表取締役 社長 田中邦裕

平素よりさくらインターネットに格別のご愛顧を賜り、誠にありがとうございます。

2011 年 12 月中旬から断続的に発生しておりました、さくらのクラウド・ストレージ障害につきまして、サービスをご利用の皆様に深くお詫び申し上げます。

現行ストレージの障害内容の詳細と新ストレージの提供に関しまして、以下の通りご報告させていただきます。

1. 障害の概要と原因、その対処

2011 年 12 月 9 日の最初の障害から、2012 年 3 月末まで断続的に障害が発生しておりました。個々の障害については発生の都度報告をさせていただいておりましたが、以下に原因ごとに整理し、その概要と原因、再発防止のための対処について説明いたします。

1.1. ストレージ装置とホストサーバ間の接続が切れる問題

サービス提供開始後、12 月中旬よりホストサーバとストレージ間のトラフィックが増加しました。これに伴い、両者のネットワーク接続が切れてしまう問題が発生しました。

1.1.1. ネットワークインターフェースの問題

ストレージ装置のネットワークインターフェースは、故障の発生に対処するため 2 つのポートを備えて冗長構成を取っています。常用系の異常は ICMP（いわゆる ping）によって監視し、これが一定期間途絶えると故障とみなし、予備系へ切り替えます。
12 月初旬より、ネットワークインターフェースにおいて断続的にパケットロスが発生し、この監視機構において一定期間の ICMP 途絶と判定される事態が発生しました。このためストレージ装置は常用系から予備系へインターフェースの切り替えを行ったのですが、切り替えの実施を原因としてユーザのサーバを収容するホストとストレージ装置間の通信が 5 分間程度切れてしまう障害が発生しました。
ただし、その後の調査により実際にはネットワークインターフェースは故障しておらず、アクセス量の増大に伴いパケットロスを引き起こしていたことが判明しました。これを回避するため、2 つの対処を行いました。

1. 1月初旬にネットワークインターフェースの監視をICMPからLINK状態の確認方式に切り替えた。これにより物理的にリンクダウンした際にのみ、予備系に切り替わるようになった
2. ストレージ装置のファームウェア・アップデートを3月19日に実施。これによって装置のパケットロスが発生しなくなった

1番目の対処を実施して以降、ネットワークインターフェースにおける冗長構成の問題は解消しました。この時点ではパケットロスの発生を完全に解消できておりませんでしたが、3月19日のファームウェア・アップデートにより解消いたしました。

1.1.2. InfiniBand ポート（Subnet Manager）の問題

12月下旬より、ネットワークインターフェースの問題の他にも、ストレージ装置へのアクセスが途絶する障害が数回発生しました。これはストレージ装置においてネットワークインターフェースが応答しなくなり、ネットワークが切れてしまうことが原因でした。
ストレージ装置を接続するネットワークはInfiniBandを用いています。InfiniBandネットワークでは、装置間の接続を管理するSM（Subnet Manager）という装置が必要となります。ネットワーク断の原因は、SMからの存在確認にネットワークインターフェースが応答しなくなり、ネットワークから装置が切り離されたと判断するために生じていました。ネットワークインターフェースには物理的な異常が見られなかったため、1月初旬にSMの確認方法を通常のモードでは不十分と判断し、heavy sweepモードに変更し、問題を解消しました。

1.2. 共有ファイルシステム数増加に伴うパフォーマンスの問題

1月初旬より、ユーザの増加にともなってストレージ上で作成される共有ファイルシステム（ユーザが使用するディスク）数が徐々に増加しました。この増加とともに、CLI（Command Line Interface）の応答が悪くなるという問題が発生しました。
CLIは、ストレージ装置を操作する際にコマンドを受け付けるインターフェースです。ストレージ装置上でファイルの作成や削除を行うために、クラウド・システムはCLIを通してコマンドを送信します。
1月頃より、CLIへの接続にかかる時間が大きくなる問題が発生しました。これを必要最小限の数にとどめるため、使われなくなったファイルシステムを頻繁に削除する必要が生じました。
しかしながら、CLIの応答と共にコマンドの処理にも同様に大きな時間がかかるようになり、クラウド・システムで実装していた手順では正常にファイルシステムの作成や削除が行えない事態が発生しました。

1.2.1. 共有ファイルシステムの作成、削除が遅くなる

ユーザのディスク作成や削除の指示は、すべてストレージ上の共有ファイルシステムの作成、削除を伴います。ファイルシステムの作成の際には、起動時間をごく短くするために、ストレージのクローン機能、スナップショット機能を利用します。
ファイルシステム数増加にともなって CLI の応答が悪くなると、これらのコマンドの実行にも非常に長い時間がかかったり、システム上タイムアウトになって正常に完了しなかったりといった問題が発生しました。これらは根本的な解決が難しく、テンプレート機能の利用停止など、ユーザの皆様に多大なご不便をおかけすることとなりました。
なお、3 月 19 日に実施したファームウェアのアップデートにおいても、当問題は解消できておらず、ファイルシステム数を一定以下に抑えるよう利用されるディスク数を抑制することで顕在化を防いでいる状況です。

1.2.2. ファイルの誤削除

ファイルシステム数が増加することがストレージのパフォーマンスの低下を引き起こす理由の一つであることから、1 月 5 日の緊急メンテナンスにおいて、負荷によって作成が正常に完了されていないディスクや、解約済みディスクの削除作業を開始しました。
しかしながら、ストレージの負荷が非常に高い状況であることから、通常の手順では削除を実行することができず、本メンテナンス専用のバッチコマンドを作成し、不要なファイルシステムの一斉削除を実行しました。
この際、レビューやテストが不十分であったことから、削除すべきファイルシステムの種類（作成に失敗したものや、解約されたものなど）が誤っていることを発見できず、稼働中のディスクの一部（53 件）を削除するという重大な事故を引き起こしてしまいました。
誤って削除されたディスクについては、ストレージのパフォーマンス低下を防ぐために、あらかじめストックとして作成していたファイルシステムであり、これらが正常に作成されていないディスクと誤認識されたことが原因でした。
なお、バックアップを取得する仕組みは用意されていましたが、ストレージの負荷が高くバックアップ頻度が低下していたことや、該当するディスクは作成されてから日が浅くバックアップが開始される前であったことから、バックアップからの復元も行えない状況となりました。
この事故の後、不要なファイルシステムの削除プロセスを単純化させるとともに、レビューやテストが十分でないスクリプトは実行させないよう徹底を行うなど、再発防止策を制定しております。

1.3. アクセス増大に伴うパフォーマンスの問題

2 月以降、ユーザの利用が本格化するに伴ってストレージへのアクセスが増大しました。こ

れに伴い、別の問題が顕在化いたしました。

1.3.1. ディスク I/O 処理の問題

ストレージのアクセスが増え、データの読み書きが頻繁になると、徐々にストレージの性能の上限に近づいていきます。処理可能な最大 IOPS に到達すると、性能が劣化し期待した処理能力を下回ってしまうことが判明しました。

この問題はストレージ装置の I/O 処理において、利用するバッファの数やプロセスの数、さらにはカーネルパラメータ等の内部状態に深く起因するものであり、メーカーとの調整の上でパラメータの変更の実施、さらにはファームウェア・アップデートを 3 月 19 日に行いました。

これらの対処により問題の一部を解消しているものの、設計仕様として期待する性能には至っておりません。

1.3.2. 監視ツールの問題

アクセスの増大に伴って、CLI のみならず管理用ツールを提供する Web インターフェースにおいても、表示が極端に遅くなる、アクセスできなくなる等の問題が発生しました。さらには、ストレージ装置の状態（コントローラの CPU、メモリの状態から、ネットワーク利用状況、ディスク I/O、各種プロトコルごとの統計情報、搭載している HDD のアクセス頻度など）を記録するログ・システムが、正しい値を取得、保存できない状態となりました。

本来、これらのツールはトラブルシュートのために必須のものですが、性能改善のためにこのツールを利用することができない状況となり、運用に多大な影響を及ぼすこととなりました。

1.4. ファイルコピー機能の動作に伴う問題

3 月 19 日にファームウェア・アップデートを実施しましたが、本来短時間の断で完了するはずの作業が長時間に渡って復旧できなくなる障害が発生しました。これは、ストレージ装置が備えるファイルコピー機能が、意図しないタイミングで実行されることにより引き起こされていました。

ファイルコピー機能は、元ファイルからコピー先ファイルへデータをコピーします。クラウド・システムはストレージ装置に対して CLI 経由でコマンドを実行しますが、これはコピープロセスを直接起動するのではなく、新しく作成したファイルに対して元ファイルからデータをコピーすることを指示する属性を与えることで実装されていました。通常であれば、この属性のチェックとコピーの実行は速やかに行われますが、ストレージ装置にかかる負荷が非常に大きいときの CLI の不具合により、クラウド・システムはコマンド実行に失敗してコピーはキャンセルされたものと解釈していました。

しかし、実際には新ファイルに属性が付与されたまま保存されていることが後に判明しました。3 月 19 日の長時間に渡るストレージ装置のブロックは、このような状況にあったファイルのコピー属性が、システムのリブートをきっかけにコピープロセスによって拾われ、実行されたことに起因していました。さらに、元ファイルは数週間前に削除されていたにもかかわらずコピーが起動されたため、ストレージ装置はディスク I/O を含め全機能がブロックするという状況に陥っていました。

3 月 29 日にも、この属性フラグとコピー機能の不整合が発生し、数十分に渡るアクセス障害の原因となりました。この時点で、上記のような問題の原因がすべて明らかとなり、全ファイルシステムの精査を行いコピー属性が残っていないことを確認して、対処を完了いたしました。

2. 現在の運用状況について

システムの安定を図るため、ストレージにかかる負荷を低減させる必要があると判断し、さくらのクラウドでは 2 台目のストレージ装置を 3 月 12 日に追加いたしました。すでにユーザの皆様にはご案内しておりますが、2 台目のストレージ装置を選択いただけるようにし、移行が可能なサーバについては 2 台目に移行していただくことで、負荷の低減、分散を進めています。

2 台目のストレージは 3 月 19 日に適用した新ファームウェアが運用開始当初から搭載されており、またファイルコピー機能は利用していないため、6 月 25 日の時点まで障害は発生していません。また 1 台目のストレージについても、負荷の低減が進むにつれて動作の安定が図られ、4 月以降ディスクの接続断となる障害は発生しておりません。

3. ストレージ装置の変更について

以上ご説明しました障害対応および運用の状況を踏まえ、現行ストレージをまったく別の新ストレージ装置に変更することといたしました。

3.1. 現行ストレージについて

さくらのクラウドを開発するにあたり、ストレージの選定の条件は以下の通りでした。

① 設計仕様に基づいたサーバ数を収容し、ディスクアクセスを処理できること
② ファイルのコピーが高速に行えること。具体的にはクローン、スナップショットの機能を有し、サーバの作成に当たって短時間で環境を提供できること
③ 高帯域アクセスが利用できること。具体的には InfiniBand ポートを備え、ネイティブで接続可能であること
④ 監視・運用に必要な機能が備わっていること
⑤ メーカーのサポートが充実しており、責任を持って運用できること

これらの条件を満たすストレージを選定し、社内テスト、β版サービス公開を経て、2011年 11 月 15 日に正式リリースいたしました。しかしながら現在までの運用において、以下のことが判明いたしました。

i. 性能限界におけるストレージ装置のテストを十分に行うことができなかった。設計仕様に基づいた最大収容数に相当するサーバを準備することができず、予想に基づいて生成した負荷により性能確認をしたため、実運用時に発生した問題に迅速かつ正確に対処することができなかった
ii. クローン、スナップショットの作成所要時間が、共有ファイルシステム数の増大により遅延することが判明した
iii. InfiniBand ポートにおいてパケットロスが観測され、期待していた性能が発揮されなかった。ファームウェア・アップデートにより解消されたが、対応完了まで 3 カ月を要した
iv. 監視・運用に必要なツール類が、ファイル数の増大、アクセスの増大に伴い利用できなくなった。このためストレージの状態を正確に把握することができなくなり、運用上重大な支障をきたすようになった
v. ストレージ装置の仕様と動作について弊社エンジニアが全容を把握することができず、発生した障害に対して十分な対応を実施することができなかった。メーカーとの綿密な連携により対処を急いだが、対処のための調査と確認に長い日数がかかってしまう結果となった

弊社の対応としましては､障害が顕在化した 12 月末以降､サービスの安定化を最優先とし、原因の究明ならびに設定変更、機材強化の実施を行うこととしていましたが、最終的にはアクセス帯域の制限による性能の限定、一部機能の停止、および新規ユーザ募集の停止という手段を取らざるを得ず、ユーザの皆様に多大なご迷惑をお掛けすることとなってしまいました。また機材強化や機材数の増加による性能向上を検討いたしましたが、設計仕様として望まれる安定性を確保することは困難であるとの判断に至りました。
以上の結果については、弊社におけるストレージ装置の選定と検証プロセスに大きな問題があったという深い反省に立ち、見直しを実施することといたしました。

3.2. 新ストレージ装置について

弊社では代替となる装置の検討を 2012 年 3 月から行い、安定性と性能の確保、および責任ある運用を実施するために、自社開発となるストレージ装置への転換を実施することとしました。
新ストレージの開発においては、上記の問題点に対し以下のように対処をしています。

I. ストレージ装置の構成は、想定収容数を確実に処理できる仕様に変更した。ストレージ装置の最大収容数に相当するテスト環境を用意し、実際に負荷をかけてテストを実施した。新ストレージの正式（課金）提供に先立ち、再度βテスト（無償提供）期間を設け、ユーザの皆様に確実に満足いただけるまでテストを繰り返すこととした
II. ディスク I/O が過大になってもコマンド処理に支障が出ないようストレージ装置あたりのディスク収容数を見直した。クローン、スナップショット機能はコピー機能で置き換えた
III. ネットワークインターフェースドライバ等は、現在までの運用実績において、問題ないホストサーバと同等のものを採用した
IV. 監視機能は、ストレージ装置上では最小限の実装とし、ディスク I/O に依存しないようにした。解析ツールはストレージ装置の外部におき、影響を受けないように変更した
V. 弊社自身で管理、運用ができるよう、自社エンジニアによる開発を行った

新ストレージ装置では、その性能限界における動作の確認に加え、QoS 設定機能も付加いたしました。これにより過大な負荷がかかった場合にも、システムを安定して運用できるよう調整することが可能となっています。
新ストレージ装置は、その機能、性能をユーザの皆様に確かめていただくため、6 月 25 日よりβテスト版（無料提供）として公開し、現在ご利用中のユーザの皆様に提供いたします。βテスト（無料提供）期間中に、十分な性能の試験、とくに性能限界においても安定性が確保できているか、運用上支障がないか、ユーザの皆様に安心して使っていただくことができるかを確認する予定です。
新ストレージ装置について十分な性能が確保できたことが確証できた後、旧ストレージ装置からの移行とともに新規ユーザ募集を再開させていただきます。

3.3. 対応の遅れについて

2011 年 12 月の最初の障害発生以降、当報告書の発表が大変遅くなったことについてお詫び申し上げます。
ストレージ装置に関する問題が顕在化して以降、弊社では現行ストレージの改善を図るべく上記の通り対応を続けてまいりました。
さくらのクラウドにおいては、ディスクに関連する機能の多くが今回問題を引き起こしたストレージの機能に依存しており、ストレージを変更するという選択肢よりもファームウェアのアップデートを持って諸問題を解決するという選択肢を優先して、解消を目指しておりました。

しかし、ストレージがブラックボックス化されていることから問題の原因が解明できず、報告書が発表できない状況が続くこととなりました。
また、3 月以降に開始した新ストレージの開発についても、信頼性の確保のためのテストなどが長期化することとなりました。

4. スケジュール

サービス正常化に向け、今後は以下のスケジュールにて進行して参ります。

4.1. 新ストレージのβテスト版（6 月 25 日）

新ストレージのβテスト版の提供を、2012 年 6 月 25 日より提供いたします。ディスクサイズについては、運用上支障がないか確認をさせていただくため、当初は 20GB に限定をさせていただきます。

4.2. 新ストレージのβテスト版の容量拡大（8 月中旬）

6 月 25 日に提供させていただく新ストレージについて 8 月まで安定性を確認させていただき、問題がなければ、より大きなディスク容量を提供させていただきます。

4.3. 新ストレージの正式運用（9 月以降）

8 月に提供を予定している、容量の拡大した新ストレージの安定性を確認させていただき、十分に課金できる品質であると判断させていただきました後に、新ストレージの正式運用を行います。

なお、新ストレージの容量拡大、正式運用の詳細な日程や新規ユーザの募集方法については、順次、弊社の Web ページ（ http://cloud-news.sakura.ad.jp/ ）にて発表させていただきます。

5. おわりに

最後になりましたが、ユーザの皆様に対し、あらためてお詫び申し上げます。
本障害のような事象を二度と発生させないために、サービス全体の信頼性を向上させ、一日も早いサービス正常化を目指して参ります。

今後とも、さくらインターネットをどうぞよろしくお願いいたします。

以上