

HPCI 共用ストレージにおける 遠隔地間メタデータサーバのフェイルオーバー

原田 浩¹ 建部 修見² 蛭原 純³ 笠原 基之⁴ 中 誠一郎³ 三浦 信一⁵ 平川 学¹

概要：HPCI 共用ストレージが採用している Gfarm2.6 では運用中のメタデータサーバのフェイルオーバー機能が追加された。HPCI 共用ストレージにおいてもメタデータサーバのフェイルオーバー導入し可用性向上をはかることとした。HPCI 共用ストレージのメタデータサーバは東京大学, 理化学研究所計算科学研究機構に分散設置されたメタデータサーバ間でフェイルオーバーを行うことによって, 従来以上のサービス継続を目指す。

1. HPCI 共用ストレージ

革新的ハイパフォーマンス・コンピューティング・インフラ (以下 HPCI) では, データ共有基盤として HPCI 共用ストレージを提供している。HPCI 共用ストレージは, ネットワーク共有ファイルシステム Gfarm を採用した容量 21.5PB の高速広域ファイルシステムで, 東京大学, 東京工業大学, 理化学研究所計算科学研究機構 (以下 AICS) に分散設置されている。メタデータサーバは東京大学と AICS にそれぞれ 2 台設置され, 東京大学設置の 1 台がマスターサーバ, 他 3 台がスレーブサーバとしてメタデータを冗長保持している (図 1)。

2. Gfarm のメタデータフェイルオーバー機能

Gfarm2.6 ではメタデータサーバのフェイルオーバー機能が追加された。Gfarm2.6 のフェイルオーバーは以下の特徴をもつクライアント透明なフェイルオーバーである。

- 運用中フェイルオーバーが可能
- Gfarm API は自動的に再接続し実行を継続, クライアントにはエラーは返らない
- ファイル書込中のフェイルオーバーに関する競合の解消

3. HPCI 共用ストレージにおけるメタデータフェイルオーバーの導入

HPCI 共用ストレージにおいてもメタデータサーバの運用中フェイルオーバーを行い可用性を向上させることとした。

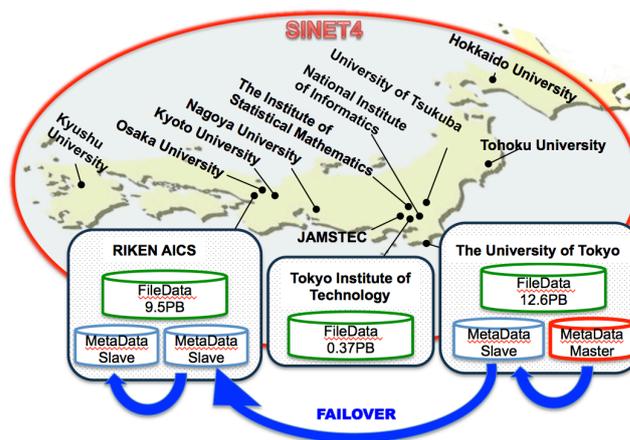


図 1 HPCI 共用ストレージにおける
メタデータサーバ間のフェイルオーバー

従来のフェイルオーバーでは, 相互監視用の専用ネットワークを用いるなどして近接設置されたサーバ間で実現されてきた。従来手法では, メタデータサーバの設置サイトにおける電源設備の法廷点検などが行われればファイルサービスを停止せざるを得ない。一方, HPCI 共用ストレージではメタデータサーバが東京大学, AICS と地理的に離れた機関に分散設置されている。分散設置されたメタデータ間でフェイルオーバーを実行することにより, 一方の設置機関においてサーバが全停止するようなメンテナンスや障害が発生してもサービス継続が可能となる。HPCI 共用ストレージでは設置機関間のメタデータサーバのフェイルオーバーを導入し, さらなるファイルサービスの継続を目指す。

¹ 理化学研究所 計算科学研究機構

² 筑波大学 計算科学研究センター

³ 東京大学 情報基盤センター

⁴ 株式会社 SRA

⁵ 東京工業大学 学術国際情報センター