

MySQL クラスタ

日本 MySQL ユーザー会
MySQL Nippon Association
<http://www.mysql.gr.jp/>

2004-11-20

MySQL クラスタ概要 1

- 歴史
 - 1991 ~ 1996 Ericson により開発
 - 2000 Alzato が設立。NDB として販売
 - 2003 年、Alzato から MySQL AB に
- MySQL 4.1 のソースに組み込まれています。
- mysql.com の宣伝文句によれば
 - 99.999% の稼働率 (停止時間は 5 分 / 年)
 - 高いスループット
- ヨーロッパのいくつかの企業では実際に業務に使用されているそうである。(スウェーデン大手のプロバイダーなど)

MySQL クラスタ概要 2

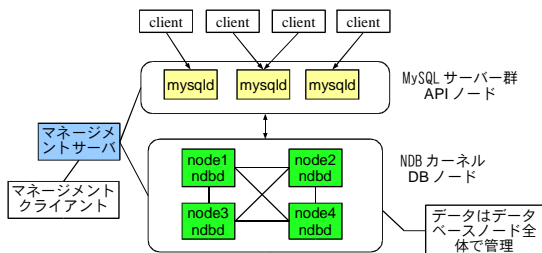
- 高速
 - 複数のノードで分散。
 - データはメモリー内にあります。
 - ログ (データベースのイメージ) は disk に記録されています
- 楽
 - 今までの MySQL との統合
 - NDB 型のテーブル (ストレージエンジン)。TYPE=ndbcluster
 - オペレータにしてみれば、他のストレージエンジンと同じ。使い勝手は同じ
 - HotBackup 機能。簡単に無停止でバックアップ
 - 障害時のノードの切り替えは一瞬。
 - ノードのメンテナンスは楽。
 - ノードを落としてまた立ち上げるだけで復旧。
- 拡張
 - mysqld や NDB のデータベースノードは、それぞれ拡張可能

MySQL クラスタ概要 3

- NDB 型は READ COMMITTED のみ。
- 行レベルロック
- shared nothing
- 各ノード間の通信方法は TCP/IP, SCI, shared-memory など複数
- 現在の対応プラットフォーム
 - Linux, Sun Solaris, IBM AIX, HP-UX, MacOS X, MS-Windows

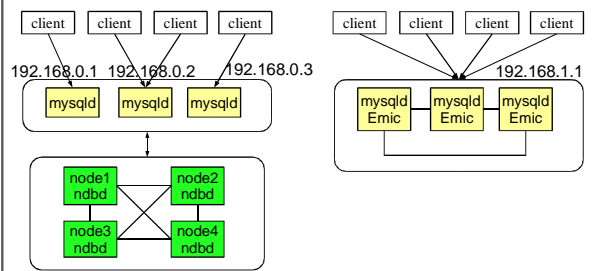
MySQL クラスタ構造 1

- 3層。それぞれの層でスケールアップ可能
- 1DB ノード = 1ndbd
- 1API ノード = 1mysqld(etc...)
- 各ノードはどのハードにインストールしてもかまわない



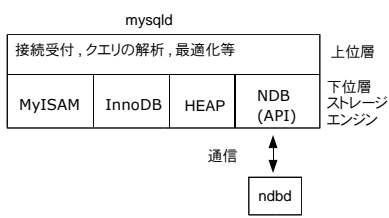
MySQL クラスタ構造 2 Emic Application Cluster との比較

- 構造、IP アドレス



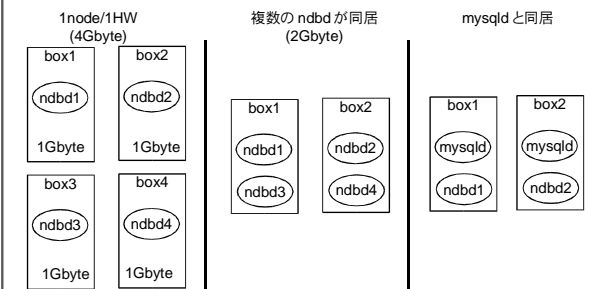
MySQL クラスタ構造 3

- mysqld にしてみれば、ndbd は別プロセス
- ndbd にしてみれば、mysqld は NDB API をもったソフト
- MyISAM, InnoDB は個々の mysqld が管理
- NDB のデータは、全ての mysqld, ndbd で共通



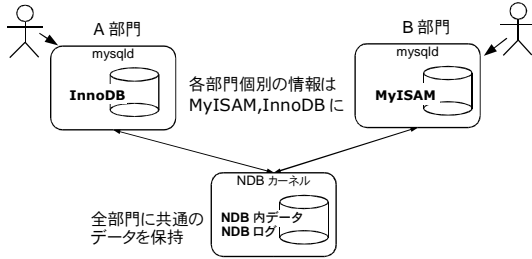
MySQL クラスタの典型的な使い方

- 1ハードウェアに1ノードにするか、1ハードウェアに複数ノードを入れるか



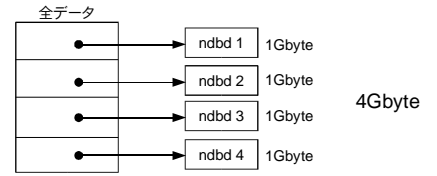
MySQL クラスターのちょっとした使い方

- たんなる冗長性、拡張性、可用性、高速性以外の目的での使い方
- 以下の特徴を逆手にとって、データアクセスの分離
 - MyISAM, InnoDB は個々の mysqld が管理
 - NDB のデータは、全ての mysqld, ndbd で共通



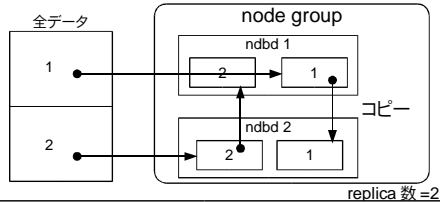
Data Fragment

- 全てのデータ (NDB 型のデータ) は、全ての DB ノードに分割されて保持管理されます。
- メリット
 - 各 ndbd に分割することで、素早い動作が期待できる。
 - ndbd 搭載のマシンを増やすだけで、使えるメモリー量 (保存できるデータ量) が増える



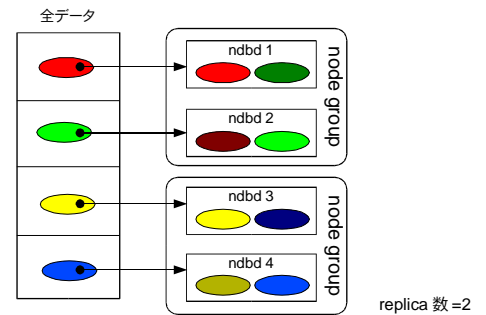
Data Replica

- ある ndbd が管理しているデータは、別の ndbd に複製されます。
 - 安全性が高まります。
 - ndbd が仮に 1 つおちたとしてもサービスは続けられます。
 - primary と secondary の切り替えは一瞬
 - ndbd のメンテナンス、障害時もサービスが提供可能になります。
- 同じデータを持っている複数の ndbd を、node group と呼びます
- コピーは同期



Data Replica と Data Fragment

- 安全性と高速性



障害検知、リカバリ

- ndbd 間はハートビートで、他が正常動作かを確認
- リカバリされたノードは、稼働中のノードからデータをコピー
- コピーが完了すればノードは完全復旧

バックアップ

- データベースのイメージログは各 ndbd が動作している機械上に作成されます。
- バックアップファイル (イメージ) も、各 ndbd が動作している機械上に作成されます。
- バックアップは、マネージメントコンソールでコマンドをたたくだけ。

現在の制限事項

- LOCK TABLES, SELECT FOR UPDATE, SELECT LOCK IN SHARE MODE は正しく動きません。
- Foreign Key は使えません。
- READ COMMITTED のみ。
- マルチバージョンニングでは無い
- 将来対応に期待

今後

- 環境の変化
 - 64bit CPU, 64bit OS, 大容量のメモリーが個人でも手に入る時代
- 要求: 大量のデータ、高いパフォーマンス
- MySQL クラスターのアプローチは期待大
- 4.1 での統合が終われば、次は 5.0 あるいは他のプラットフォームへの対応
- easy setup. 設定は mysqld とちがって必要。設定が楽になるように変更中。