データベース多重化運用時のディスク異常監視

仕組みと設計例

技術を知る

システムの冗長化を行う際には、事業の継続性を確保し、損害を最小限に抑え、迅速な復旧を可能にするための対策をしておく必要があります。データベースの観点では、データベースプロセス、ディスク、ネットワークなど、データベース運用の継続に不可欠な要素の障害を検知し、フェイルオーバーやスタンバイサーバの切替えといった対策を講じることが求められます。本記事では Fujitsu Enterprise Postgres（以降、Enterprise Postgres）のデータベース多重化運用を支える Mirroring Controller（以降、MC）機能を用いたディスクの異常監視の設定方法について解説します。ディスク障害を早期に検知し、フェイルオーバーなどの対策を実行するためのパラメーター設定について詳しく説明します。設計を行う際の参考にしてください。

ディスク異常検知の仕組み

ディスク異常検知の設定は、システム要件である目標復旧時間（RTO：Recovery Time Objective）に合わせて調整する必要があります。Enterprise Postgres における MC のディスク異常検知は次の 2 つの仕組みにて行われます。これらの監視は並行して行われ、先に異常を検知したタイミングで縮退が発生します。設定する際は 1、2 の順番で行います。

1. ディスクチェック処理が連続して異常となり、リトライ回数を超過した場合
2. ディスクチェック処理が無応答となり、そのタイムアウト時間を超過した場合

これらの設定が不適切であると、システム要件を満たせなくなる可能性があるため、仕組みを理解した上で適切に設定することが重要です。

監視の仕組み

1. ディスクチェック処理のリトライ超過による縮退

ディスクチェック処理で異常と判定された回数がリトライ回数を超えると、MC が異常を検知し縮退が行われます。設定できるパラメーターは以下のとおりです。

説明	パラメーター名	デフォルト値
異常監視の監視間隔	disk_check_interval	800（ミリ秒）
異常監視のリトライ回数	disk_check_retry	2（回）

異常と判断するまでの目安時間は下記の計算式となります。デフォルト値を使用した場合、障害発生タイミングにもよりますが、おおよその異常検出時間は 2.4 秒です。

```
disk_check_interval（ミリ秒）/ 1000 ×（ disk_check_retry（回数）+ 1 ）
```

図 1 は、障害発生後、初回のディスクチェック処理で異常が検出され、その後 2 回のリトライでも異常と判断された場合に縮退が行われている例を示しています。図中の「異常検出時間」は、障害が発生してから縮退が起こるまでの時間を指しています。

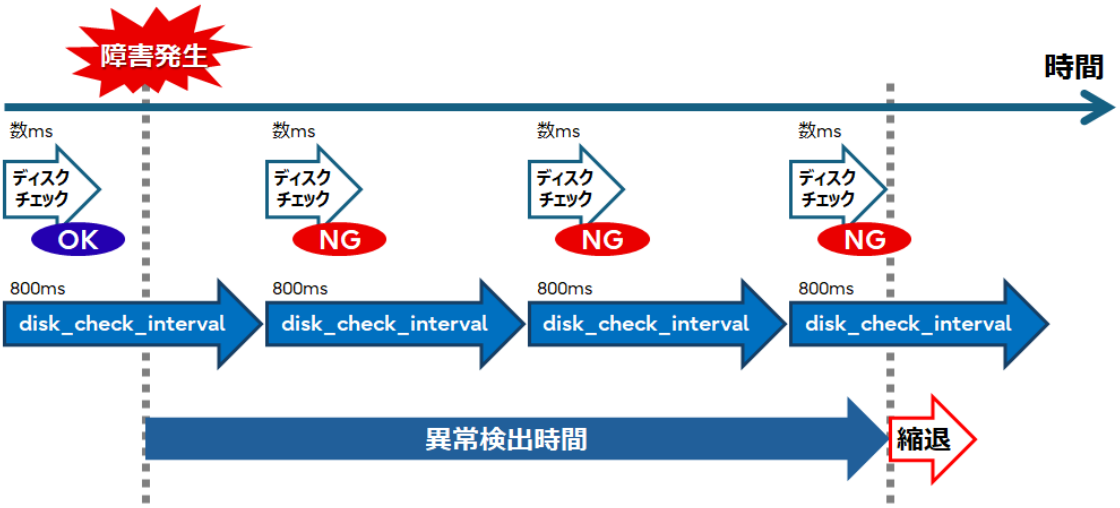

図 1：ディスクチェック処理のリトライ超過による縮退

2. タイムアウト時間の超過による縮退

ディスク異常については、ディスクが無応答になる可能性も考慮する必要があります。例えば、障害によりディスクチェック処理が応答しない場合、リトライ超過による縮退ができなくなり、要件を満たせなくなる可能性があります。そのため、ディスクチェック処理全体のタイムアウト時間を設定し、異常を検知する仕組みが必要です。MC には、ディスクチェック処理が無応答の時間が、設定したタイムアウト時間（disk_check_timeout）を超過した場合に、異常を検知し縮退を行う機能があります。設定できるパラメーターは以下のとおりです。

説明	パラメーター名	デフォルト値
異常監視のタイムアウト時間	disk_check_timeout	2147483 (秒)

図 2 は、ディスク障害発生後、最初のディスクチェック処理がパラメーターで設定したタイムアウト時間を超過したタイミングで縮退が行われている例です。
なお、「disk_check_timeout」パラメーターは、障害発生後の初回のディスクチェック開始からの時間です。そのため、図 2 の"A"の時間を考慮して、設定値を決める必要があります。この"A"の時間は最大で「disk_check_interval」となります。

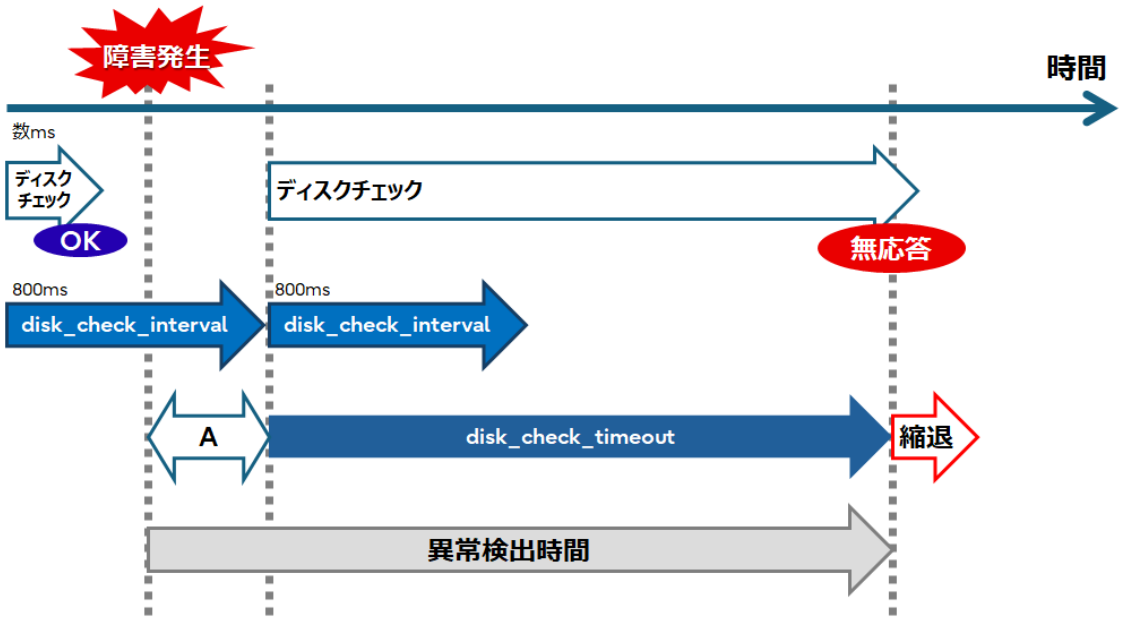

図 2：タイムアウト時間の超過による縮退

設計例

これまで述べた異常検知の仕組みをもとに、具体的なパラメーターの設計例を紹介します。ここに示す例は、データベースに障害が発生した際のシステムダウンタイムを最小限に抑えるにあたり、データベースの切り替え時間を 1 分以内に完了することを目標とし、ディスク異常検知から縮退を開始するまでの時間を 5 秒以内にするという要件があるものとします。なお、監視するディスクの数は 1 つとします。

1. ディスクチェック処理のリトライ超過による縮退

要件に合うように各パラメーターを設計します。インターバルの時間（disk_check_interval）を 900ms に設定し、リトライ回数（disk_check_retry）を 4 と設定しました。このパラメーター設定時のおおよその異常検出時間は 4.5 秒であり、要件の切り替え時間（5 秒）が満たせる設定となります。なお、インターバルの時間が短く、リトライ回数が非常に多い設定では、システムの負荷が高まる可能性があります。

2. タイムアウト時間の超過による縮退

次にタイムアウト時間（disk_check_timeout）を決めます。「障害発生からディスクチェック開始までの時間（図内の"A"）」は最大でおおよそ disk_check_interval の 0.9 秒となるため、disk_check_timeout は要件の切り替え時間（5 秒）から 0.9 秒を減算し、小数点以下を切り捨て 4 秒と設定します。

上記の設定を図示すると図 3 のようになり、どちらのメカニズムでも切り替え時間内に縮退が起こる設定になっていることが確認できます。

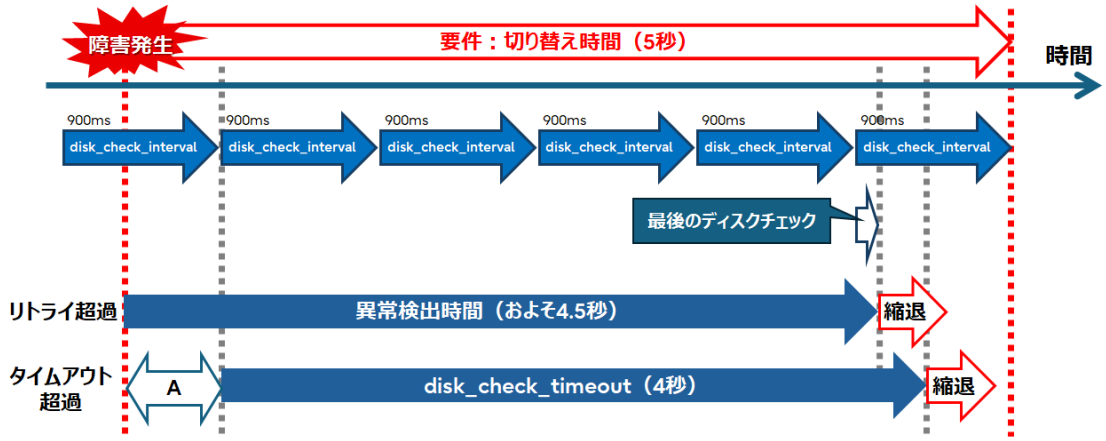

図 3：要件に沿った設定例

ポイント

ディスク異常監視の設計にて注意するべきポイントを紹介します。詳細はマニュアルを参考に設計してください。

スタンバイサーバのタイムアウト異常による切り離し設定

スタンバイサーバでタイムアウトによる異常検知時に、スタンバイサーバの切り離しを行うために shutdown_detached_synchronous_standby の設定を on にする必要があります。

コピーコマンドを使用したバックアップ / リカバリ利用時のタイムアウト時間

データをコピーするにあたり、ETERNUS のコピー機能である「QuickOPC（Quick One Point Copy）」を利用する場合、以下の処理を考慮して設計する必要があります。
アドバンスト・コピー機能によるコピー処理からファイルシステムを保護するために、データ格納先（複製元ボリューム）のファイルシステムを一時的に凍結する処理があります。 この凍結による縮退が起こらないようにするため、ディスク異常監視のタイムアウト時間（異常検出時間）は、凍結処理時間よりも大きな値に設定する必要があります。

ディスクチェックの並列処理

異常監視に利用するスレッド数の上限を決めるパラメーター（disk_check_max_threads）があります。このパラメーターを適切に設定することでディスクチェック処理を並列に行うことができます。

参考

- Enterprise Postgres クラスタ運用ガイド（データベース多重化編） ― 「2.11.4.4 ディスクの異常監視のチューニング」
- Enterprise Postgres 運用ガイド ― 「付録 L ETERNUS ディスクアレイのアドバンスト・コピー機能を利用したコピーコマンドのサンプル」

2025 年 3 月 11 日