SRA OSS, Inc. ホワイトペーパー

# PowerGres V7 ウォームスタンバイ対応機能について


2010-01-08

SRA OSS, Inc. 日本支社
〒170-0005 東京都豊島区南大塚 3-46-3 大塚セントコアビル 5F
Tel. 03-5951-1191　Fax. 03-5951-1192
http://www.sraoss.co.jp/
powergres-info@sraoss.co.jp

# 目次



# 1. 本ドキュメントの目的

PowerGres on Linux、PowerGres on Windows のバージョン 7（以下 PowerGres V7）では、付属の GUI 管理ツールに「ウォームスタンバイ対応機能」が加わりました。この機能拡張では、従来から知られていた「ウォームスタンバイ構成」を構築するのを助けるだけでなく、二つのサーバ間でのデータ通信にあたり、追加コンポーネントを必要としないという従来にない特長を備えています。

本ドキュメントでは PowerGres V7 で新たに登場したウォームスタンバイ対応機能について、使い方、メリット、アーキテクチャーを解説します。

また、PowergGres V7 / PostgreSQL 8.4 におけるウォームスタンバイの基本的な解説もあわせて行います。一般的なウォームスタンバイについて熟知している方は3章から読み進めてください。

***PowerGres とは***

PowerGres とは、代表的なオープンソースデータベースソフトウェアの PostgreSQL をベースにした、SRA OSS,Inc. が開発販売するデータベースソフトウェア製品です。GUI 管理ツールと独自の拡張が加えられている。また、PowerGres むけにサポートサービスが提供されます。

# 2. ウォームスタンバイとは

ウォームスタンバイとは、PowerGres/PostgreSQL における高可用性クラスタリング技法のひとつです。1台の稼動系サーバに対して、1台の待機系サーバを用意し、WAL ファイル（トランザクションログファイル）を転送することで、稼動系の更新内容を待機系に伝え、自動バックアップを実現します。また待機系では、起動準備状態（リカバリをしている状態）を続けることで、逐次に WAL ファイルを読み込みしておくことで、リカバリ時間をきわめて短くします。ログシッピングなどとも呼ばれます。（図1）

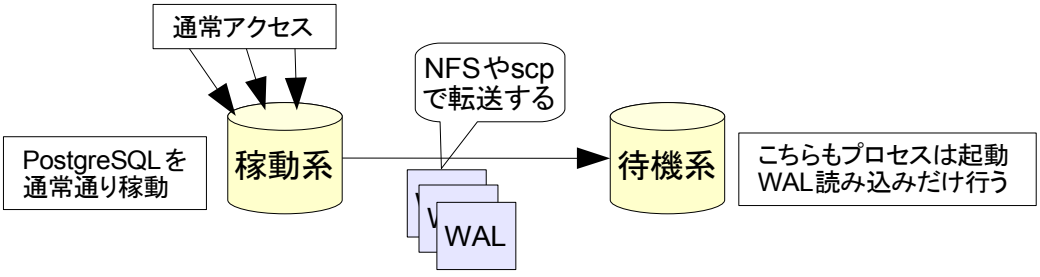



図1：ウォームスタンバイ構成

ウォームスタンバイ構成の稼動系サーバの設定は、通常の単体の PowerGres/PostgreSQL と変わりません。必要なことは、アーカイブモードを有効に設定するだけです。具体的には postgresql.conf の設定で archive_mode を on に、archive_command にファイルコピーを行うコマンドを記述します。必要に応じて、archive_timeout も設定します。[詳しくは PostgreSQL マニュアル／24.3. 継続的アーカイブとポイントインタイムリカバリ（PITR）を参照ください]

これらの点は、他のクラスタ方式、レプリケーション方式と比べると簡便です。また、アプリケーション側も、ウォームスタンバイ構成を組んでいることをまったく意識する必要はありません。単体のデータベースと同様にアクセスできます。

しかしながら、ウォームスタンバイの待機系サーバを用意するには、PowerGres/PostgreSQL 自体のほかにいくつかコンポーネントが必要となります。また、ウォームスタンバイ構成は、高可用性クラスタリングの方法としては、いくつか制限事項があります。

### *ウォームスタンバイ構成のための追加要素*

まず、（1）稼動系から待機系にファイルを転送するための手段が必要となります。scp コマンドや rcp コマンドを使う方法と、NFS やファイル共有サービス（Windows の場合）を使う方法がよく使われます。

また、どちらのサーバが主体となって転送を行うか構成方法にバリエーションがあります。方法と手順を以下の表にまとめます。いずれにせよ、PowerGres/PostgreSQL とは別にファイルを転送する仕組みを用意する必要があります。

| | | |
|---|---|---|
| scp/rcp | プッシュ型 | 稼動系の archive_command に待機系へ WAL ファイルをコピーするコマンドを書く。 |
| scp/rcp | プル型 | 待機系の何らか設定で、稼動系から WAL ファイルをコピーしてくるコマンドを書く。 |
| NFS/ファイル共有 | プッシュ型 | 待機系のアーカイブログディレクトリを共有して、稼動系の archive_command で、そこに WAL ファイルをコピーするコマンドを書く。 |
| NFS/ファイル共有 | プル型 | 稼動系のアーカイブログディレクトリを共有して、待機系の何らか設定では、そこから WAL ファイルを読み込むようにする。 |

方法の得失としては、プッシュ型では待機系が止まっている場合に稼動系にエラーが出てしまいます。実害はありません。プル型は稼動系のクラッシュで喪失する WAL ファイルが多くなり、場

合によっては待機系を素直に起動できないケースがあります。たいていはデータベースクラスタに pg_resetxlog コマンドをかければ起動できますが、データ損失の可能性が残ります。

さて、次に（2）待機系のリカバリをコントロールするソフトウェアが必要となります。通常の PITR では、WAL ファイルが無くなった時点で、リカバリ作業をやめて、データベースを起動します。明示的な指示がない限り、これを起動させないで WAL ファイルを待ち続けるようにさせるスクリプト、ソフトウェアが必要となります。

PowerGres V6、PostgreSQL 8.3 から、contrib/pg_standby というツールが付属するようになり、（2）を行う事実上の標準ソフトウェアとなりました。待機系の recovery.conf の中の設定 recovery_command に

```
'pg_standby -t /tmp/pg.trigger.5432 /mnt/warmstandby/pg_arc %f %p %r'
```

などと記述すれば、簡単に ウォームスタンバイが構成できます。この場合、pg_arc ディレクトリに稼動系の WAL ファイルが順次送られてくる前提で、リカバリを行います。また、（稼動系障害時に）/tmp/pg.trigger.5432 ファイルを置くと、待機系サーバがリカバリを中断して、起動します。

pg_standby は指定した WAL ファイルのあるディレクトリに対して、ファイル存在チェックやファイル消去などを行うため、scp を使ったプル型の構成には対応できません。［pg_standby について詳しくは PostgreSQL マニュアル／ F.26. pg_standby を参照ください］

### *ウォームスタンバイ構成の制限事項*

ウォームスタンバイ構成は、他の高可用性クラスタ構成の方法と比べると以下の制限事項があります。

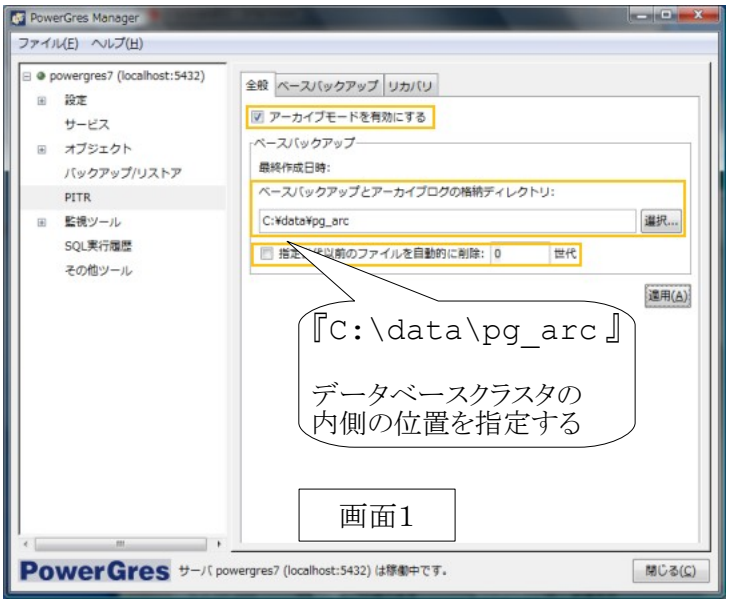
- スタンバイ側に SQL でアクセスすることはできません（参照アクセスも不可です）
  - PostgreSQL8.5 でこの制限を解消する機能が投入される予定です
- 自動フェイルオーバーの機能を持ちません
  - 「障害を検知して、待機系を起動して、アプリケーションからのサーバ接続先を何らかの方法で切り替える」という機能は別途用意が必要です
- 待機系と稼動系は同じ OS／CPU アーキテクチャのマシンでないといけません
  - 32bit、64bit の違いがあると動きませんが、それ以外の OS のエディションの違いなどは問題とならないことが多いです。データベースクラスタの物理バックアップがそのまま使用できるかで確認できます。
- 稼動系がクラッシュした場合、最後にコミット済みであるデータの幾分かが失われる可能性があります
  - archive_timeout 設定により、データが失われるリスクを定量的に限定することが可能です。10 分に設定すれば、少なくとも障害の 10 分前くらいの状態で、待機系を起動することができると考えられます。

なお、ウォームスタンバイ構成の基本原理については PostgreSQL マニュアルにも記述があります。［PostgreSQL マニュアル／24.4. 高可用性のためのウォームスタンバイを参照ください］

# 3. PowerGres V7 のウォームスタンバイ対応機能

PowerGres V7 は、二つのサーバに各々 PowerGres を導入することで、ウォームスタンバイ構成を用意に構築できるようになっています。そのために「設定 GUI」と「WAL ファイル転送手段」の二つを提供します。

### *GUI 機能と使い方手順*



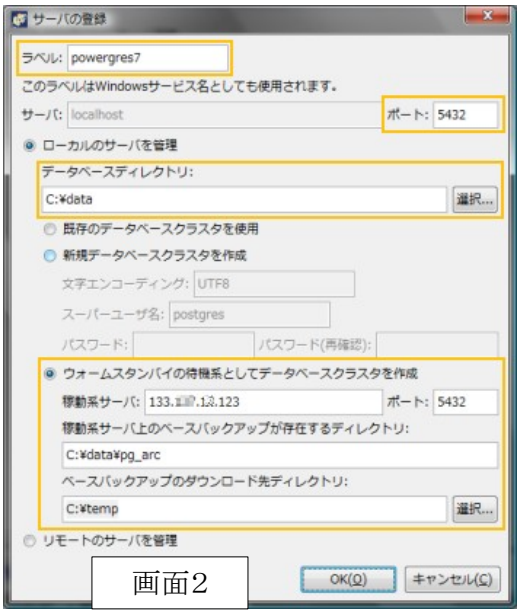

まず、稼動系の PowerGres サーバを構築します。

ツリーメニューの『PITR』を選び、設定パネルの『全般』ページでアーカイブモードを有効にします。

少し奇妙ですが『ベースバックアップとアーカイブログの格納ディレクトリ』にデータベースクラスタのディレクトリ中のディレクトリを指定します。(画面1)

ベースバックアップを取るときにここで指定したディレクトリは自動的に除外してくれますので、自分自身を何重にも含んで、ベースバックアップファイルが肥大化することはありません。

ベースバックアップを取るたびに古いアーカイブログ等を自動的に削除してくれる機能がありますが、ウォームスタンバイのときには、これは使わないようにします。



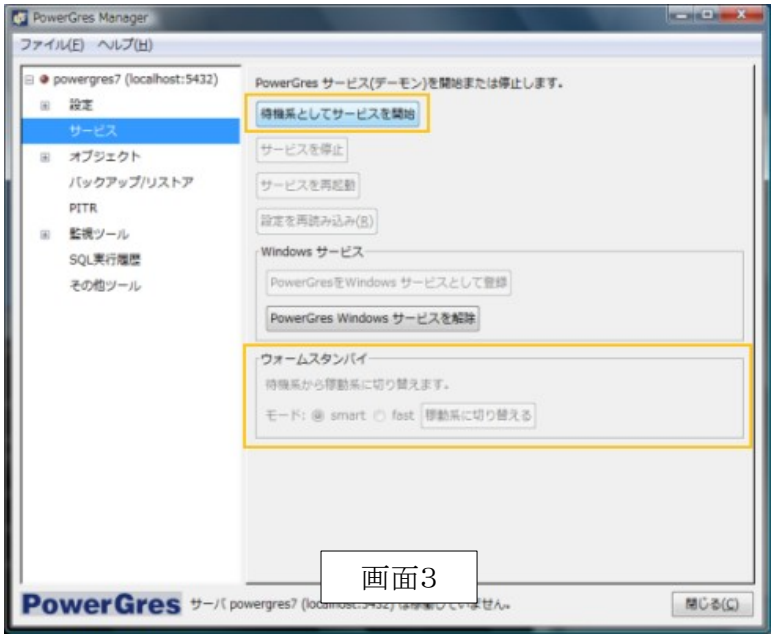

次に待機系のサーバを構築します。PowerGres の『サーバの登録』のダイアログで『ウォームスタンバイの待機系としてデータベースクラスタを作成』を選択します。(画面2)

ここで、稼動系サーバのホスト名か IP アドレス、ポート番号、稼動系サーバのベースバックアップ(およびアーカイブログ)が存在するディレクトリ、一時的なダウンロード先のディレクトリを指定します。もちろん、待機系 PowerGres 自体のデータベースディレクトリやポート番号も指定します。

これで、待機系としてサーバが登録されます。『サービス』メニューで『待機系としてサービスを開始』ボタンを押せば、ウォームスタンバイが開始されます。(画面3)

稼動系に切り替えるのも、『稼動系に切り替える』というボタンを押すことで実行できます。その際に smart モードと fast モードを選択できます。fast モードではすぐさま起動しますが、smart モードではその時点で存在する WAL ファイルを可能か限り適用してから起動します。

なお、『稼動系に切り替える』ボタンは内部的にはトリガーファイル設置で実現されています。待機系のデータベースクラスタディレクトリにファイル名が「powergres.trigger」で内容が文字列「fast」もしくは「smart」というファイルを置くことで、起動させることができます。



画面3

### *WAL ファイル転送アーキテクチャー*

PowerGres V7 では、ウォームスタンバイ構成に、データベースクライアントが接続するのと同じポートにて libpq プロトコルを使ったファイル転送を使っています。libpq プロトコルとは、PowerGres/PostgreSQL にクライアントが接続するときに使うプロトコルです。形態としては待機系からのプル型に相当しますが、待機系内にも一時保管のディレクトリを持ち、稼動系クラッシュ時の WAL 損失を抑止します。（図2）

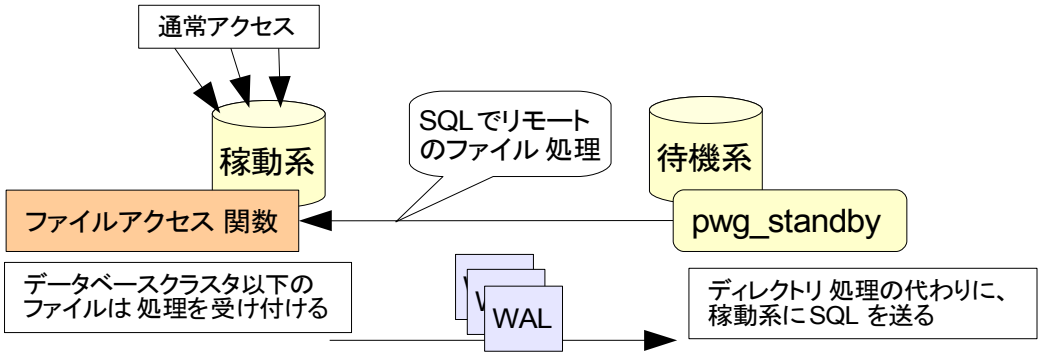


図2：PowerGres V7 のウォームスタンバイ構成

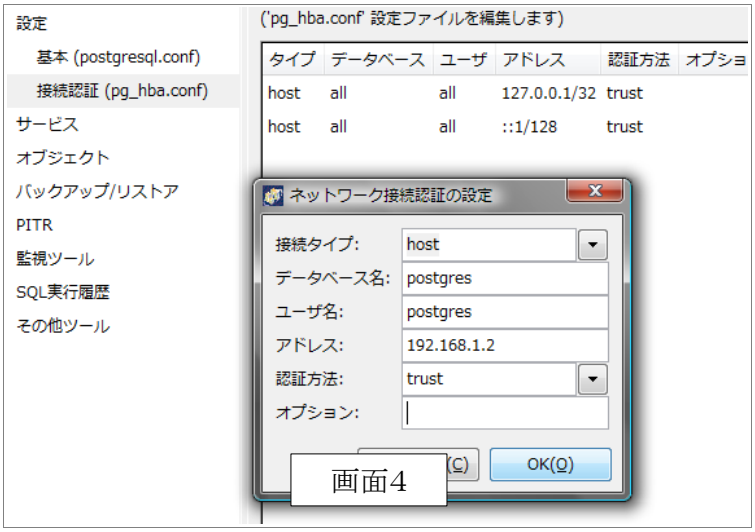
待機系には pg_standby の替わりに、pwg_standby というツールを使用します。pg_standby はディレクトリを監視して、新たな WAL ファイルがあれば、これを待機系のデータベースサーバが読む場所に移動するものですが、pwg_standby は稼動系サーバに SQL を投げて、新たな

WAL アーカイブファイルがあるかを確認して、SQL を通して WAL 自体のバイナリデータも受け取ります。

PowerGres V7 サーバにはファイルに対する処理要求を受け付ける SQL から使える関数が用意されています。これによってファイルの存在確認、ファイルの削除、ファイル内容の取得などを行います。

これら関数の大部分はベースとなる PostgreSQL8.4 にも存在しているものですが、バイナリファイルを転送できる関数 pg_read_bin_file() だけが PowerGres V7 の拡張となります。これらはデータベースクラスタ以下のファイルパスしかアクセスできないようになっているため、ベースバックアップとアーカイブログ（WAL ファイル）の格納場所が、データベースクラスタ内に限定されます。また、これらの関数は管理者権限をもったユーザ（ロール）でしか使用できないようになっているため、ウォームスタンバイ用にも管理者ユーザを使用します。[ファイルアクセス関数について詳しくは PostgreSQL マニュアル 9.24 システム管理関数（汎用ファイルアクセス関数）を参照]

一般に管理者ユーザからの接続は、サーバ稼動ホスト内からしか許さない設定にすることが多いですが、それに加えて、ウォームスタンバイのペアとなるマシンに対しても許容するように pg_hba.conf ファイルを設定する必要があります。この設定も PowerGres の GUI から行うことができます。（画面4）



画面4

## *メリットと活用方法*

ウォームスタンバイは簡易にできる可用性向上策として前バージョンのころから使われてきました。PowerGres V7 ウォームスタンバイ対応機能は、これをさらに手軽に簡単にしたといえます。メリットとして以下のような点が挙げられます。

- GUI だけで設定、待機開始、待機から稼動への切り替え、ができます。
- ファイル転送について Linux、Windows とも共通の SQL アクセス（libpq プロトコル）を使った方法を提供します。OS の機能を駆使する必要がありません。NFS や CIFS、鍵認証による scp などの設定を行う必要がありません。
- 使用するポートが PowerGres 自体と共用であるためウォームスタンバイ構成にするため、ほとんどの場合、ネットワーク設定を変える必要がありません。

主要な活用目的としては、**『日次バックアップやアーカイブロギングより、もう一段階の信頼性・可用性向上を狙う場合の一手として導入する』**ということが考えられます。