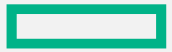


Hewlett Packard
Enterprise

PostgreSQL プロセス障害の話

Noriyoshi Shinoda

February 2, 2021

SPEAKER
篠田典良（しのだのりよし）

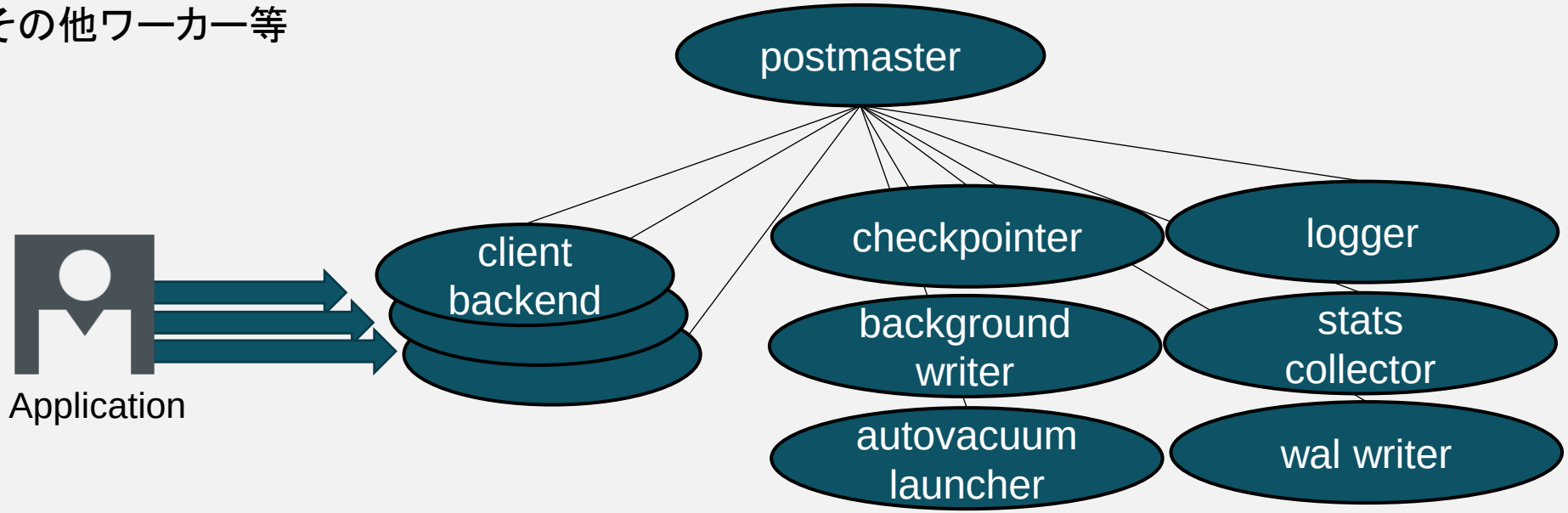
- 所属
  - 日本ヒューレット・パッカード株式会社
- 現在の業務
  - PostgreSQL をはじめ、Oracle Database, Microsoft SQL Server, Vertica 等 RDBMS 全般に関するシステムの設計、移行、チューニング、コンサルティング
  - Oracle ACE（2009 年 4 月～）
  - Oracle Database 関連書籍15冊の執筆
  - オープンソース製品に関する調査、検証
- 関連する URL
  - 「PostgreSQL 虎の巻」シリーズ
    - http://h30507.www3.hp.com/t5/user/viewprofilepage/user-id/838802
  - Oracle ACE ってどんな人？
    - http://www.oracle.com/technetwork/jp/database/articles/vivadeveloper/index-1838335-ja.html

最近の活動
PostgreSQL 14 向け

- CREATE OR REPLACE TRIGGER 文に対する psql コマンドの TAB 補完
  - Committed
- pg_stat_replication_slot カタログの列名変更
  - Committed
- オンライン整合性チェック関数（pg_relation_check_pages）
  - バグ・レポートを提出
  - Revert !
- 暗号化のための鍵管理システム
  - バグ・フィックス・パッチを提出（Committed）
  - Revert !!
- psql コマンドから拡張統計の一覧を出力する（¥dX）
  - バグ・レポートを提出
  - Revert !!!　⇒ Committed

プロセス障害の話
インスタンスのプロセス構成

- Postmaster を親とするプロセス群
  - クライアント・バックエンド
  - 必須バックエンド・プロセス
  - その他ワーカー等



プロセス障害の話
プロセス障害時の挙動

- restart_after_crash パラメーター(デフォルト on)により変化する
- PostgreSQL 12 のマニュアル

> デフォルトであるonの場合、PostgreSQLはバックエンドのクラッシュの後、自動的に再初期化を行います。この値を真のままにしておくことが、通常データベースの可用性を最大化する最適の方法です。しかし、PostgreSQLがクラスタウェアにより起動された時のような状況では、クラスタウェアが制御を獲得して、適切とみなすいかなる振る舞いをも行えるように再起動を無効にすることが有益かもしれません。

- PostgreSQL 13 のマニュアル

> When set to on, which is the default, PostgreSQL will automatically reinitialize after a backend crash. Leaving this value set to on is normally the best way to maximize the availability of the database. However, in some circumstances, such as when PostgreSQL is being invoked by clusterware, it may be useful to disable the restart so that the clusterware can gain control and take any actions it deems appropriate.

プロセス障害の話
プロセス障害時の挙動

- 「バックエンドのクラッシュ」とは？
  - KILL シグナルで強制終了
  - Segmentation Fault (SEGV) 等による異常終了
  - 「src/backend/postmaster/postmaster.c」の reaper() で処理

- restart_after_crash = on （デフォルト）の場合
  - 異常終了したプロセスだけでなく、関連する全プロセスが再起動される
  - 全クライアント・セッションがリセットされ、全トランザクションがロールバックされる

- restart_after_crash = off の場合
  - Postmaster プロセスを含めインスタンス全体が停止する

プロセス障害の話
プロセス障害時の挙動

– 再初期化後のプロセスID例

プロセス	停止前PID	停止	再起動後PID	備考
postmaster	38786		38786	
logger	38787		38787	
checkpointer	38789		38908	
background writer	38790		38909	
walwriter	38791	kill –9	38910	
autovacuum launcher	38792		38911	
archiver	38793		38912	
stats collector	38794		38913	
logical replication launcher	38795		38914	
postgres	38830		-	client backend

プロセス障害の話
プロセス障害時の挙動

– PostgreSQL 12 のマニュアル
  – パート VI. リファレンス / PostgreSQLサーバアプリケーション / postgres

> -n
> これはサーバプロセスが異常終了する問題をデバッグするためのオプションです。 このような状態では、他のすべてのサーバプロセスに対し、 終了し、 共有メモリやセマフォを再初期化するように通知するのが、 通常の戦略です。 これはエラーが起きたサーバプロセスが、 終了する前に、 何かしら共有状態を破損したかもしれないからです。 このオプションは、 postgresが共有データ構造を再初期化しないように指定します。 知識のあるシステムプログラマであれば、 その後にデバッガを使用して共有メモリやセマフォの状態を検証することができます。

プロセス障害の話
プロセス障害時の挙動

- 要求に応じて起動されるので再起動されない
  - クライアント・バックエンド
  - autovacuum worker
  - autoprewarm worker
  - logical replication worker
  - parallel worker
  - walsender

プロセス障害の話
プロセス障害時の挙動

- 異常終了すると restart_after_crash を無視して再起動するプロセス
  - logger
  - stats collector
  - archiver

⇒ 全プロセスの再初期化は行われない

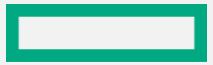
- 異常終了すると restart_after_crash を無視してインスタンス全体が停止するプロセス
  - postmaster
  - startup

プロセス障害の話
クライアント・バックエンドの停止

– プロセス操作関数

関数名	説明	送信シグナル	備考
pg_cancel_backend(*pid*)	SQL文の停止	INT	
pg_terminate_backend(*pid*)	セッションの停止	TERM	
pg_reload_conf()	構成ファイルの再ロード	HUP	Postmaster ⇒ 全バックエンドへ
pg_rotate_log()	ログ・ローテーション	USR1	Postmaster ⇒ Loggerへ

– Pid は pg_stat_activity カタログの pid 列から参照



プロセス障害の話
クライアント・バックエンドの停止

- KILL シグナルを送ると
  - Parallel Worker を強制終了した例

```
postgres=> EXPLAIN ANALYZE SELECT COUNT(*) FROM data1 a, data1 b;
WARNING:  terminating connection because of crash of another server process
DETAIL:  The postmaster has commanded this server process to roll back the current transaction and exit,
because another server process exited abnormally and possibly corrupted shared memory.
HINT:  In a moment you should be able to reconnect to the database and repeat your command.
server closed the connection unexpectedly
        This probably means the server terminated abnormally
        before or while processing the request.
The connection to the server was lost. Attempting reset: Failed.
!?>
```

- クライアント・バックエンドや Parallel Worker も restart_after_crash の影響を受ける
- OOM Killer 等による単一プロセス停止が全体に波及する

プロセス障害の話
archive_command の障害

- archive_command に指定されたコマンドの障害
  - src/backend/postmaster/pgarch.c　内の pgarch_archiveXlog 関数で実行
  - pgarch_archiveXlog 関数が false を返すと1秒待って再実行

```
pgarch_archiveXlog(void)
{
    rc = system(xlogarchcmd);
    if (rc != 0)
    {
        if (WIFEXITED(rc))
        {
            ereport(lev, (errmsg("archive command failed with exit code %d" …
        } …

        return false;
    }
```

THANK YOU

Mail: noriyoshi.shinoda@hpe.com
Twitter: @nori_shinoda