

システム構成

UI作成、ファイルアップロード	index.php
UTF8変換、PostgreSQL取り込み	dbinsert.py

簡単なHTMLによるファイルアップロードフォームをPHPで作成しました。

ファイルのアップロードが終わった時点で、PHPから dbinsert.py が呼び出されます。

dbinsert.py では、文字コードをUTF8に変換後、psycopg2モジュールを使って、PostgreSQLのテーブルにデータを挿入しています。

テスト

お預かりしたテストデータ（ハローページsample1009）を1000回コピーして、1,008,000件のCSVファイルを作成しました。

http://teleapolist.info/postgres_test/index.php にアクセスし、フォームからCSVデータをアップロードしました。

ファイルのアップロードにかかる時間は、利用者の端末の性能やネットワークの状況により変わります。

変換及び取り込みの時間は、365～385秒／100万件 でした。

操作方法など

アップロードフォームのURL：<http://teleapolist.info/apitest/index.php>

CSVファイルアップロード

テーブルの初期化をする

チェックすると、テーブルを初期化してからデータを取り込みます。

CSVファイルを分割してアップロードする場合、2つ目以降はチェックをしないようにします。

CSVファイル: sample.csv

このボタンを押して、送信するCSVファイルを選択します。

ファイルを送信します。

送信後、自動でUTF8変換とPostgreSQLへの取り込みをします。

画面の表示が変わるまで、そのままお待ちください。

データ確認

簡単なデータ確認用のツールです。

データ確認ツールのURL：<http://teleapolist.info/apitest/dbdump.php>

URLの後ろに、?s=100&c=50 とすると、101件目から50件を表示します。（例：<http://teleapolist.info/apitest/dbdump.php?s=100&c=50>）

（デフォルトは、s=0 c=100）

仕様

アップロードできるファイルのサイズ

お預かりしたテストデータを1000回コピーしたファイルのサイズを元に、256Mをサイズの上限としました。
.htaccess ファイルでPHPの設定を変更しています。

データベースへの接続

ユーザーアカウント	teleapolist	PW:pdbm66
rootアカウント	root	PW:aptakknksg
PostgreSQLアカウント	postgres	PW:aptakknksg
PostgreSQL DBユーザー	postgres	(PW:なし)
データベース名	teleapolist	(PW:なし)
テーブル名	teleapolist	

テーブルの構成 (UTF-8)

key	integer 連番 <i>primary key</i>	テーブル内で一意となるキー。(挿入時に自動付加)	
tel	character varying(13)	電話番号	
name	character varying(255)	会社名	
zipcode	character varying(8)	郵便番号	
pref	character varying(32)	都道府県名	
adrs1	character varying(255)	住所 1	
adrs2	character varying(255)	住所 2	
code	character varying(255)	?	
industry	character varying(255)	業種	
s_tel	character varying(13) index	検索用電話番号	2015.03.05 追加
s_name	character varying(255) index	検索用会社名	2015.03.05 追加

s_tel: 0から9までの数字のみ格納

s_name: 半角文字を全角文字に変換して格納

履歴

2015.03.02 root でのログインを確認しました。
2015.03.03 サーバーに vsftpd をインストールしました。
2015.03.03 /home/teleapolist/public_html/postgres_test ディレクトリを作成しました。
2015.03.03 サーバーに python-psycopg2 をインストールしました。
2015.03.03 teleapolist のパスワードを pbdm66 に変更しました。
2015.03.03 ftp 転送テスト。
2015.03.04 /usr/lib/python/site-packages/ に、sitecustomize.py を設置しました。
2015.03.04 PHPの設定を変更するために、.htaccessファイルを設置しました。
2015.03.04 index.php、dbinsert.py を設置し、1,008,000件のデータのアップロード及び、DB取り込みのテストをしました。
2015.03.04 簡単なデータ確認用ツールを作成しました。
2015.03.05 Python-2.7.4 をビルドしました。
2015.03.05 zenhan.py をインストールしました。
2015.03.06 検索用フィールドを追加 (s_tel, s_name)
s_tel: 電話番号のうち、ハイフンを除く数字のみを格納
s_name: 半角文字を全角文字に変換して格納
tel、name フィールドの INDEX を、s_tel、s_name に変更
2015.03.08 pgpool II を導入 (/usr/local/bin/pgpool)
2015.03.08 pgpool テスト
2015.03.08 memcached 導入
2015.03.08 pgpool + memcached テスト
2015.03.12 レスポンスの内容を変更
[responseHeader][countALL] を追加 (start、rows を無視して検索した全件数)
[response][doc][要素番号] で、要素番号を PostgreSQLの格納番号に変更
2015.03.13 レスポンスの内容を変更
[response][doc][要素番号] で、要素番号を 検索結果中の出現番号に変更
2015.03.14 ディレクトリ作成、スクリプト移動
http://teleapolist.info/
apist/
 .htaccess
 hello01.php API本体
apitest/
 .htaccess
 index.php データベースメンテナンス (データ入替え、追加スクリプト)
 dbinsert.py データベースメンテナンス (index.php から呼び出される Pythonスクリプト)
 api_sample.php PHPからAPI (hello01.php) を利用するサンプルプログラム
 app_tools.php アクセスキーの発行、カウンタ確認をするスクリプト
 dbdump.php データベースに登録されたデータをチェックするためのスクリプト
2015.03.15 ドキュメント改訂