

ОПТИМИЗАЦИЯ СТРУКТУРЫ ДАННЫХ АРХИВА АСУ ТП АЭС В ФОРМАТЕ POSTGRESQL С ТОЧКИ ЗРЕНИЯ ОБЪЕМА ХРАНЕНИЯ И СКОРОСТИ ВЫПОЛНЕНИЯ ЗАПРОСОВ ДЛЯ ПРОГРАММНОГО КОНВЕРТЕРА ИЗ БИНАРНЫХ ФОРМАТОВ АРХИВОВ

Степанов В. Н.

Институт проблем управления им. В.А. Трапезникова РАН,
Россия, г. Москва ул. Профсоюзная д.65
vnstepanov@yandex.ru

Аннотация: В работе описываются подходы к оптимизации структуры данных архива АСУ ТП АЭС и их реализация в программе-конвертере из двух бинарных форматов архивов, разработанных в ИПУ РАН в формат PostgreSQL.

Ключевые слова: АСУ ТП, АЭС, архив, PostgreSQL, структура данных.

Введение

Архив АСУ ТП АЭС [1-3] представляет собой журнал истории всех сигналов, регистрируемых сервером. Он содержит информацию о виде и типе сигнала, его значении, времени возникновения сигнала и признаке достоверности значения и времени. Поток сигналов может быть очень большим, поэтому алгоритм записи и структура архива оптимизированны с точки зрения скорости записи, в ущерб возможностям последующего поиска и анализа данных. Исходный формат архивов представляет собой набор бинарных файлов, каждый из которых содержит все изменения значений сигналов за одну минуту. Некоторые задачи анализа данных еще можно выполнять с такой структурой данных, например при расследовании нештатных ситуаций, время возникновения которых известно. Но для статистической обработки и анализа данных такой формат не подходит. С другой стороны задачи статистического анализа накопленных за период времени сигналов не требуют быстродействия реального времени, в отличие от задачи записи данных в архив. Для конвертации архивов сигналов в формат, удобный для статистической обработки, был разработан программный конвертер из бинарного формата в формат СУБД PostgreSQL [4,5].

1 Структура данных архива в формате PostgreSQL

Структура создаваемой БД показана на Рис.1.

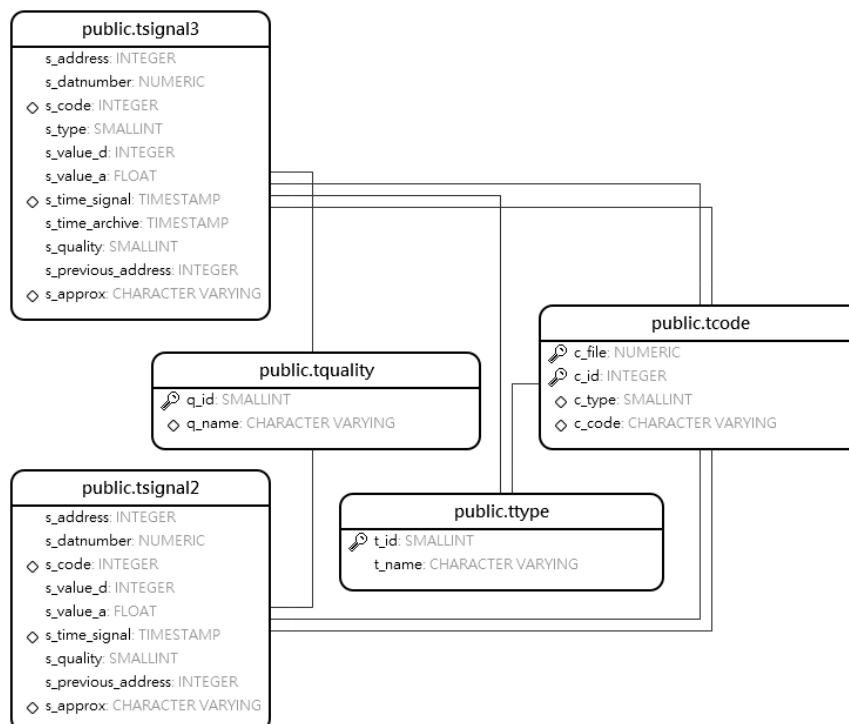


Рис. 1. Универсальная структура базы данных архива в формате PostgreSQL

Эта структура основана на минимально необходимом наборе полей и минимально возможном размере каждого поля, для обеспечения отсутствия потери информации при конвертации из двух бинарных форматов архивов АСУ ТП АЭС. Минимизация необходима, т.к. количество записей в БД очень большое. Реакторный отсек 2го энергоблока АЭС Куданкулам генерирует за 2 месяца около 3млрд сигналов. И если, например, для поля код сигнала (s_code) таблицы tsignal2 использовать тип данных int8 вместо необходимого и достаточного int4, объем БД увеличится на 1 Гб .

Все связи условны, каждая таблица автономна, целостность данных не поддерживается и уникальность не обеспечивается. Подразумевается, что исходный архив не содержит ошибочной и дублированной информации, а дополнительные проверки при вставке каждой записи непреемлемо увеличивают время импорта. Кроме того, это позволяет избежать возможных ошибок, например при обнаружении типа данных, не включенного в таблицу ttype. Таблицы ttype и tquality являются справочными и не участвуют в запросах. Основная таблица сигналов создается сразу в двух форматах, для старого и нового форматов архива. Это сделано для минимизации размера получаемой БД. Новый формат (таблица tsignal3) содержит больше полей, которые не использовались бы в случае импорта туда архивов старого формата.

Для ускорения выполнения запросов в таблицах tsignal2, tsignal3 создаются индексы по полям s_code,s_time_signal,s_approx.

Таблица 1. Описание полей таблиц архива

tsignal2	s_address	Int4	Адрес (смещение) сигнала в исходном файле архива
	s_datnumber	Int8	Число-имя связанного *.dat файла
	s_code	Int4	Код сигнала из *.dat файла
	s_value_d	Int4	Значение дискретного сигнала (если сигнал аналоговый – null)
	s_value_a	Float4	Значение аналогового сигнала (если сигнал дискретный – null)
	s_time_signal	timestamp	Время прихода сигнала
	s_quality	Int2	Код достоверности сигнала
	s_previous_address	Int4	Ссылка на предыдущее значение сигнала (искать по полю s_address)
s_approx	boolean	Флаг аппроксимированного значения (сигнал из периодической части)	
tsignal3	s_address	Int4	Адрес (смещение) сигнала в исходном файле архива
	s_datnumber	Int8	Число-имя связанного *.dat файла
	s_code	Int4	Код сигнала из *.dat файла
	s_type	Int2	Тип сигнала
	s_value_d	Int4	Значение дискретного сигнала (если сигнал аналоговый – null)
	s_value_a	Float4	Значение аналогового сигнала (если сигнал дискретный – null)
	s_time_signal	timestamp	Время прихода сигнала
	s_time_archive	timestamp	Время записи сигнала в архив
	s_quality	Int2	Код достоверности сигнала
	s_previous_address	Int4	Ссылка на предыдущее значение сигнала (искать по полю s_address)
s_approx	boolean	Флаг аппроксимированного значения (сигнал из периодической части)	
tcode	c_file	Int8	Число - Имя файла (без расширения)
	c_id	Int4	Код сигнала
	c_type	Int2	Тип сигнала
	c_code	varchar	Технологический индекс сигнала
ttype	t_id	Int2	Номер типа
	t_name	varchar	Описание типа
tquality	q_id	Int2	Код качества сигнала
	q_name	varchar	Описание кода

2 Описание программного инструмента конвертации из бинарных форматов архива в формат PostgreSQL

Программа конвертации архивов в БД PostgreSQL состоит из следующих файлов:

Имя файла	Описание
Arc2sql	Основной исполняемый модуль
arc2sql_partitioned_code.sql	Sql скрипт создания структуры БД с секционированием таблицы сигналов по коду сигнала
arc2sql_partitioned_time.sql	Sql скрипт создания структуры БД с секционированием таблицы сигналов по времени сигнала
arc2sql_plain.sql	Sql скрипт создания структуры БД со сплошной таблицей сигналов, без поддержки секционирования

Основной исполняемый модуль использует следующие параметры командной строки (символом «*» отмечены необходимые параметры, указание которых обязательно):

- -i - путь к исходному архиву для конвертации (по-умолчанию архив ищется в текущем каталоге)
- -o - путь к временной папке для служебных промежуточных файлов (по-умолчанию – текущий каталог)
- -H - * ip адрес сервера PostgreSQL
- -u - имя пользователя для подключения к серверу PostgreSQL (по-умолчанию «postgres»). Пароль для этого пользователя должен быть предварительно задан в глобальной переменной среды PGPASSWORD. Удобно задавать эту переменную в bash скрипте и из него же вызывать исполняемый модуль.
- -d - * имя базы данных для импорта архива.
- -m - Тип основной таблицы с сигналами:
 - 0 – единая сплошная таблица (по-умолчанию). Подходит в тех случаях, когда минимизация размера БД приоритетнее скорости выполнения запросов к ней.
 - 1 – таблица сигналов секционирована по времени. Подходит в тех случаях, когда большинство запросов планируется с условием выборки по временным диапазонам. Используется вложенное секционирование, Сначала таблица разбивается на 2 секции – основные сигналы и периодические. Затем секция основных сигналов разбивается на секции по времени сигнала. Диапазон одной секции – 1 час. Секция периодических сигналов дополнительно на подсекции не разбивается, т.к. в большинстве запросов эти сигналы не участвуют.
 - 2 – таблица сигналов секционированна по коду сигнала. Подходит для тех случаев, когда большинство запросов планируется с выборками по коду сигнала. Секционирование производится аналогично типу 1, но секция основных сигналов разбивается на секции по коду сигнала с диапазоном 1 код.
- -l - уровень подробности информации, выводимой в файл журнала в процессе конвертации. 0 – выводить только ошибки. (по умолчанию - 999 -максимум информации). Файл журнала создается в папке, указанной в параметре -o.
- -r - уровень подробности информации, выводимой в консоль в процессе конвертации. (по умолчанию - 0 -выводить только ошибки).
- -h - вывод справочной информации по параметрам командной строки.

SQL Скрипты вызываются из исполняемого модуля в зависимости от параметра «-m».

Основной исполняемый модуль выполняет последовательно следующие действия:

- Создание БД с именем из параметра «-d» на сервере «-h».
- Заполнение структуры БД путем вызова соответствующего параметру «-m» SQL скрипта
- Автоматическое определение формата архива (старый или новый) по типам файлов с данными. Если в каталоге «-i» содержатся файлы с расширениями *.ia, *.id, *.idv, принимается старый формат архива. Если файл *.ar – новый формат.
- Поиск файлов с расширением *.dat в каталоге «-i» и импорт их в таблицу tcode базы данных «-d».
- Поиск файлов с расширениями *.ia, *.id, *.idv, *.ar в каталоге «-i» и импорт их в таблицу tsignal2 или tsignal3. Импорт (п.4, п.5) происходит следующим образом:
 - Файл архива конвертируется в файл специального бинарного формата с расширением *.bin, поддерживаемого PostgreSQL сервером
 - С помощью штатной командной консоли PostgreSQL (psql) файл *.bin импортируется в

- таблицу tsignal2 или tsignal3 в зависимости от формата исходного архива
- Файл *.bin удаляется.

Последовательный режим импорта (п. 5) вместо пакетного выбран с целью минимизации требуемого для импорта свободного места на носителе. *.bin файлы занимают существенно больше места, чем исходные файлы архивов, т.к. содержат кроме самих данных много служебной информации, такой как размер полей.

Заключение

Описанный программный инструмент конвертации бинарных архивов двух базовых форматов в формат СУБД PostgreSQL позволяет создавать структуру базы данных с тремя типами секционирования основной таблицы сигналов, что позволяет оптимизировать скорость выполнения запросов для разных задач статистической обработки и анализа данных. Секционирование по времени подходит для запросов статистики по всем сигналам (или определенного типа) в узких временных диапазонах. Секционирование по коду сигнала подходит для анализа изменения значений определенных сигналов.

Литература

1. Менгазетдинов Н.Э., Бывайков М.Е., Зуенков М.А., Промыслов В.Г., Полетыкин А.Г. и др. Комплекс работ по созданию первой управляющей системы верхнего блочного уровня АСУ ТП для АЭС "Бушер" на основе отечественных информационных технологий [Электронный ресурс]: монография. М.: ИПУ РАН, 2013. – ISBN 978-5-91450-130-0.
2. АЭС «КУДАНКУЛАМ» Блоки 3,4. Автоматизированная система управления технологическими процессами. Программное обеспечение сервера. Структура архивов. М.: ИПУ РАН, 2019. R188.KK34.0.0.AP.RP.WD003.460.0229530.00003-06 90 01
3. АЭС «КУДАНКУЛАМ». Блоки 3,4. Автоматизированная система управления технологическими процессами. Программное обеспечение сервера. Структура архивов. М.: ИПУ РАН, 2020. R188.KK34.0.0.AP.RP.WD003.460.0229530.00003-06 90 03-ЛУ.
4. Васильев А. Ю. Работа с PostgreSQL: настройка и масштабирование. Creative Commons Attribution - Noncommercial 4.0 International 2017
5. PostgreSQL 10.21 Documentation. The PostgreSQL Global Development Group. Copyright © 1996-2022 The PostgreSQL Global Development Group