

УДК 004.652.4

UDC 004.652.4

**К ВОПРОСУ СБОРА ДАННЫХ
ЭЛЕКТРОЭНЕРГЕТИЧЕСКИХ СИСТЕМ**

**THE ISSUE OF DATA COLLECTION IN
ELECTRICITY SYSTEMS**

Атрощенко Валерий Александрович
д.т.н., профессор

Atroschenko Valery Aleksandrovich
Dr.Sci.Tech., professor

Фишер Антон Владиславович
аспирант

Fisher Anton Vladislavovich
postgraduate student

Дьяченко Роман Александрович
к.т.н
*ФГБОУ ВПО «Кубанский государственный
технологический университет», Краснодар, Россия*

Dyachenko Roman Aleksandrovich
Cand.Tech.Sci.
*Kuban State Technological University, Krasnodar,
Russia*

В статье рассматриваются два способа организации масштабирования системы хранения хронологических данных. Приведены формулы расчета оптимальной продолжительности периода (для разбиения данных по периодам), а также предельно допустимой частоты поступления данных при которой гарантируется их сохранение системой

The article discusses the two ways to scaling storage chronological data systems. The formulas for calculating the optimal duration of the period (to split the data by period) and the maximum allowable rate for guaranteed preservation in the system are presented

Ключевые слова: БАЗА ДАННЫХ, НАСЛЕДОВАНИЕ, МАСШТАБИРОВАНИЕ, ПАРТИЦИРОВАНИЕ, СУБД, ХРОНОЛОГИЧЕСКИЕ ДАННЫЕ, ЭЛЕКТРОЭНЕРГЕТИКА

Keywords: DATABASE, INHERITANCE, SCALING, PARTITIONING, DBMS, CHRONOLOGICAL DATA, ELECTRICITY

Введение

В процессе проектирования распределенных информационных систем сбора информации, мониторинга и прогнозирования в электроэнергетике возникает проблема организации хранения хронологических данных полученных от отдельных узлов системы. А также проблема обеспечения способности к масштабированию всей системы в целом на достаточном уровне. Под масштабированием информационных систем сбора и хранения данных понимается возможность добавления новых узлов в рабочую систему с допустимо малыми изменениями в самой системе.

Общая структура систем сбора хронологических данных

Рассмотрим распределенную систему подачи электроэнергии с N территориально распределенными узлами. На рисунке 1 изображены потоки информации в системе сбора данных.

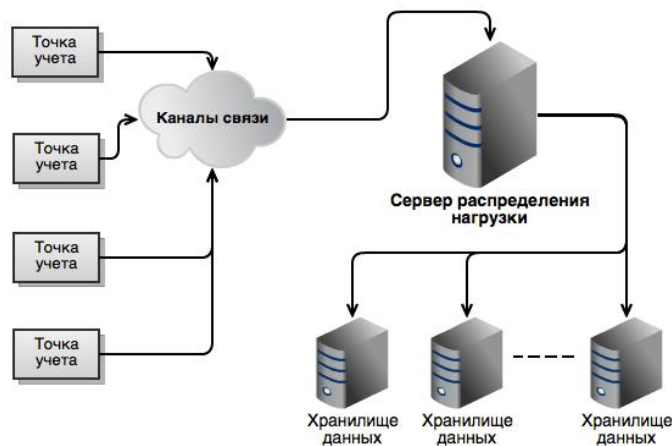


Рисунок 1. Потоки информации в системе сбора данных

Задачей ставится разработка такой структуры базы данных, которая, во-первых, позволит хранить данные, полученные от различных узлов системы, таким образом, чтобы средняя скорость выполнения следующих запросов выбора наборов данных была наивысшей:

- выбор данных потребления за последние сутки (неделю, месяц, квартал, год) по всем точкам учета (ТУ);
- выбор данных потребления за последние сутки (неделю, месяц, квартал, год) по конкретной ТУ.

Во вторых структура позволит добавлять новые точки учета с наименьшими затратами по времени простоя и настройки измененной системы, а также наращивать дополнительные ресурсы без структурных изменений центральных узлов системы. То есть позволит создать максимально масштабируемую систему.

Очевидно, что использование для решения поставленной задачи схем баз данных, использующих одну таблицу для хранения всех поступающих от точек учета данных, не позволит лучшим образом решить поставленную задачу обеспечения масштабируемости. Так как при использовании подобных решений постоянно растущий объем таблицы в определенный момент создаст известные трудности в работе серверов баз данных из-за существующих ограничений оперативной памяти (ОЗУ) вычислительных систем. Скорость выборки данных при запросе к базе данных зависит от объема ОЗУ вычислительной машины доступной серверу базы данных и непосредственно размера самой таблицы к которой производится запрос. График зависимости времени выполнения запроса выборки данных от количества записей в таблице представлен на рисунке 2.

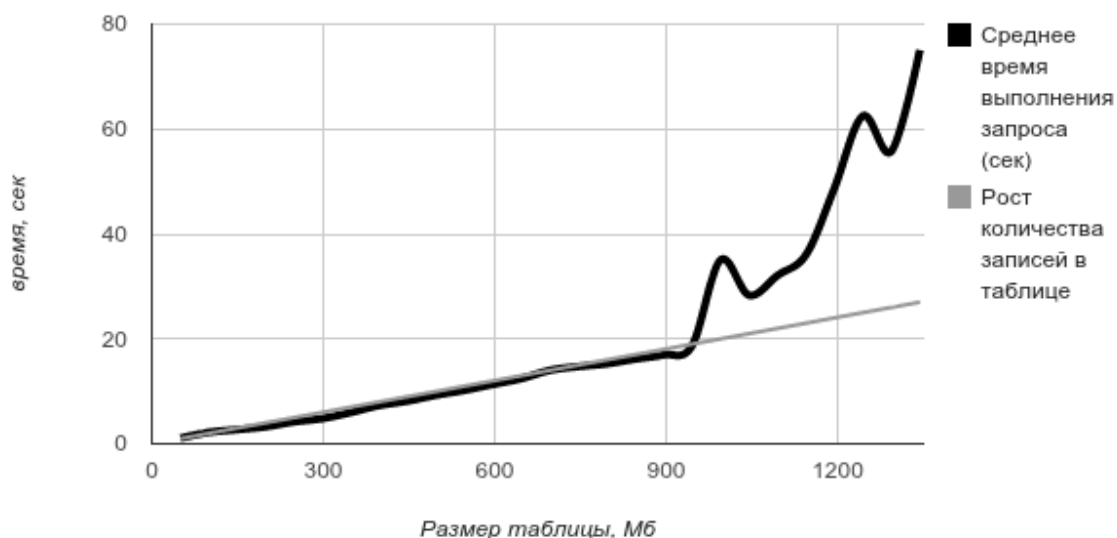


Рисунок 2. Зависимость времени выполнения запроса от размера таблицы.

Замеры производились на вычислительной машине с доступными 1024 Мб оперативной памяти для СУБД. Как видно при достижении замером таблицы размера ОЗУ выделенного для работы СУБД время выполнения

запросов начинает увеличиваться не линейно. Следовательно максимальная производительность запросов в СУБД возможна при условии, когда суммарный размер таблиц (с учетом размера дополнительных данных необходимых для выполнения запроса) участвующих в запросе меньше размера ОЗУ доступного СУБД.

Решением проблемы хранения хронологических данных может послужить разбиение поступающей информации на части на основании определенного критерия. Как минимум, возможно использование одного из двух критериев. Первый предусматривает разбиение данных по самим источникам их поступления, а второй – разбиение данных по времени их поступления. Решение проблемы масштабируемости системы в данном случае ложится на компонент системы отвечающий за распределение нагрузки (вычислительная машина принимающая все входные данные).

Способ с разбиением данных по точкам учета

Первым способом решения проблемы хранения хронологических данных поступающих от различных источников является выделение отдельных таблиц непосредственно на ТУ. Недостатки этого метода хорошо видны при попытке построения запроса вычисляющего среднее значение собираемой величины (в случае системы учета электроэнергии собираемая величина – объем потребления энергии) по всем ТУ. Чтобы выполнение таких запросов было возможным необходимо модифицировать структуру хранилища данных (ХД).

При использовании технологии наследования таблиц, имеем структуру ХД, изображенную на рисунке 3.

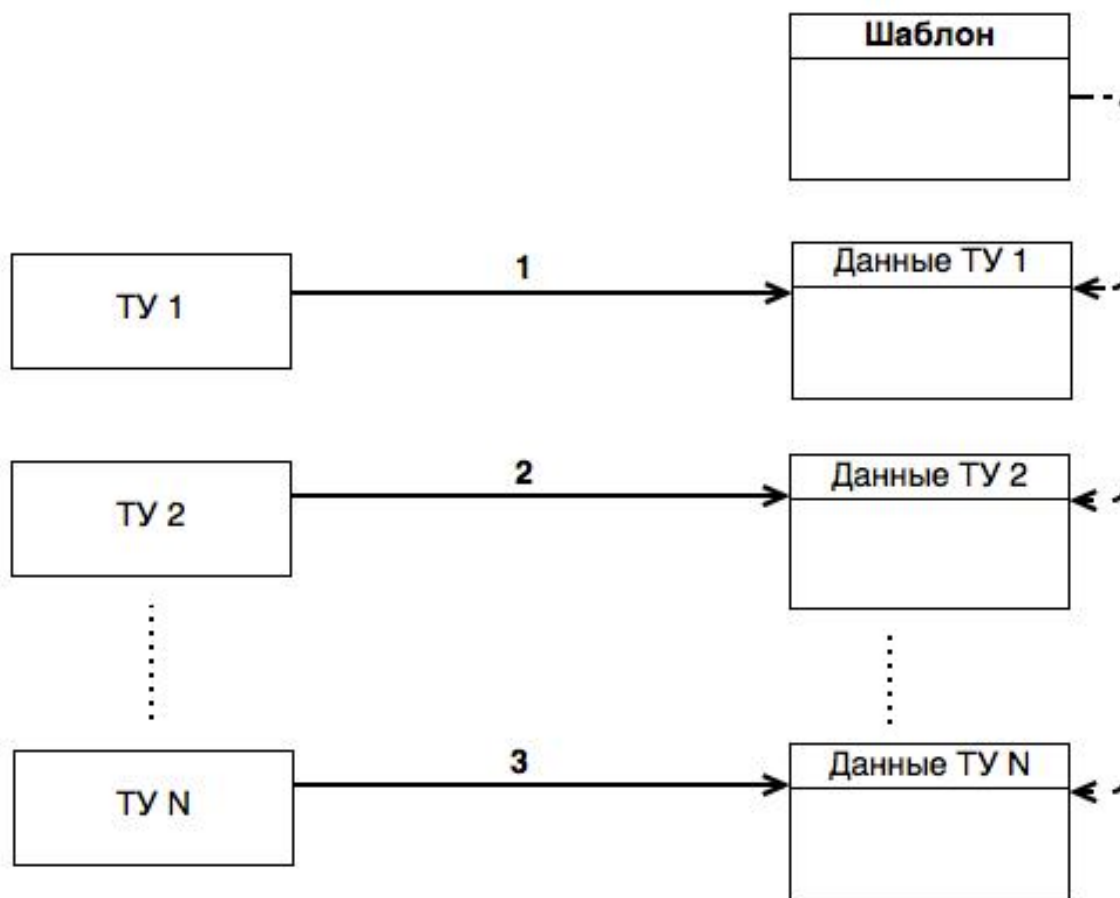


Рисунок 3. Разбиение данных по ТУ с наследованием таблиц

Данные от ТУ 1-*N* поступают соответственно в таблицы 1-*N* созданные по общему шаблону. Количество таблиц равно количеству точек учета. Таблицы имеют общий шаблон – это позволяет выполнять сквозные запросы обрабатывающие данные по всем таблицам.

Приведенный метод разбиения данных решает проблему выборки средних значений, но не решает проблему переполнения таблиц данными. При необходимости внесения в систему нового узла учета необходимо добавить и новую таблицу, что допустимо, но не достаточно удобно в плане масштабируемости. Данная методика успешно применима в случае хранения не большого объема данных, а также в случае проблем с передачей данных

центральному узлу обработки (отсутствие подходящего канала связи). В таких случаях каждая ТУ собирает свои данные, храня их в единообразной форме, определяемой центральным узлом обработки данных. Данные синхронизируются при появлении соответствующей возможности. Предельно допустимая интенсивность поступления данных от N ТУ, которые должны быть сохранены, определяется по формуле:

$$\lambda_{\text{пред}} = \frac{M}{N \cdot m \cdot t}, \left[\frac{\text{зап}}{\text{с}} \right] \quad (1)$$

где M – доступный объем оперативной памяти вычислительной машины (которая может быть использована СУБД для выполнения операций над данными) [Мб],

N – количество точек учета (количество таблиц),

m – объем памяти занимаемый одной записью $\left[\frac{\text{б}}{\text{зап}} \right]$,

t – время в течении которого необходимо хранить данные [с].

Значение предельно допустимой интенсивности поступления данных численно равно частоте опроса f , выраженной в Гц. Следовательно период опроса T составит:

$$T = \frac{1}{f}, [\text{с}] \quad (2)$$

В случае хранения данных за период 1 год (или ~ 31556926 с) для 100 ТУ, при доступном объеме оперативной памяти $M = 4294967296$ б и размере одной записи в 50 б, допустимая интенсивность поступления данных равна:

$$\lambda_{\text{пред}} = \frac{4294967296}{100 \cdot 50 \cdot 31556926} \approx 0,02722 \left(\frac{\text{зап}}{\text{с}} \right).$$

Период опроса равен:

$$T = \frac{1}{0,02722} \approx 36,73769 \text{ (с)}$$

В случае масштабирования системы построенной на данном принципе, т.е. при добавлении в рабочую систему дополнительной ТУ, возрастет интенсивность поступления записей. Так как первоначальный расчет предельно допустимой интенсивности поступления записей теперь не позволит предотвратить превышение допустимого объема оперативной памяти для работы системы, необходимо либо увеличить объем доступной памяти, либо выполнить перерасчет допустимой интенсивности с учетом новой ТУ. Вариант решения проблемы путем увеличения памяти, является не достаточной удобным и оперативным. Так как эта операция подразумевает остановку работы системы, подбор, установку и настройку нового оборудования. Даже в случае использования систем позволяющих наращивать доступные мощности динамически (“налету“) имеется некоторое предельное ограничение. Второй вариант использующий перерасчет предельно допустимой интенсивности поступления записей является более универсальным и не требует дополнительных действий со стороны оператора или дополнительного обслуживания системы. Суть метода заключается в том, что после перерасчета допустимой суммарной интенсивности, элемент системы отвечающий за распределение нагрузки (вычислительная машина, сервер) в соответствии с новым значением ограничивает поступления данных в базу отсекая часть поступающих данных, но при этом гарантируя сохранения производительности системы в целом. С увеличением точек учета линейно уменьшается количество сохраненных данных одной ТУ. Следует

отметить, что этот способ не лишает систему необходимости выполнения работы по созданию дополнительной таблицы для новой ТУ.

Способ с разбиением по времени поступления

Решение проблемы хранения хронологических данных поступающих от различных источников путем разбиения данных по времени поступления позволяет выполнять запросы, касающиеся выборки значений потребления энергии за определенные периоды времени, с меньшими затратами ресурсов, чем без использования разбиения. Выбор оптимальных периодов для разбиения данных основывается на условиях будущих запросов к базе данных. Получая прирост производительности выборок на небольших периодах (при разбиении данных на более мелкие части) система потребует больше времени на выборки по большому диапазону и наоборот. В рассматриваемом случае возможно использование следующих периодов: сутки, неделя, месяц, квартал и год. При разбиении на периоды необходимо учитывать количество данных, которое необходимо будет хранить. В общем случае, объем данных, хранящийся в наиболее часто используемом для выборок данных периоде, не должен превышать объем оперативной памяти вычислительной машины доступный серверу базы данных. Если запас памяти велик, то делить данные на более мелкие части не имеет большого смысла. На рисунке 4 изображено распределения данных по периодам.

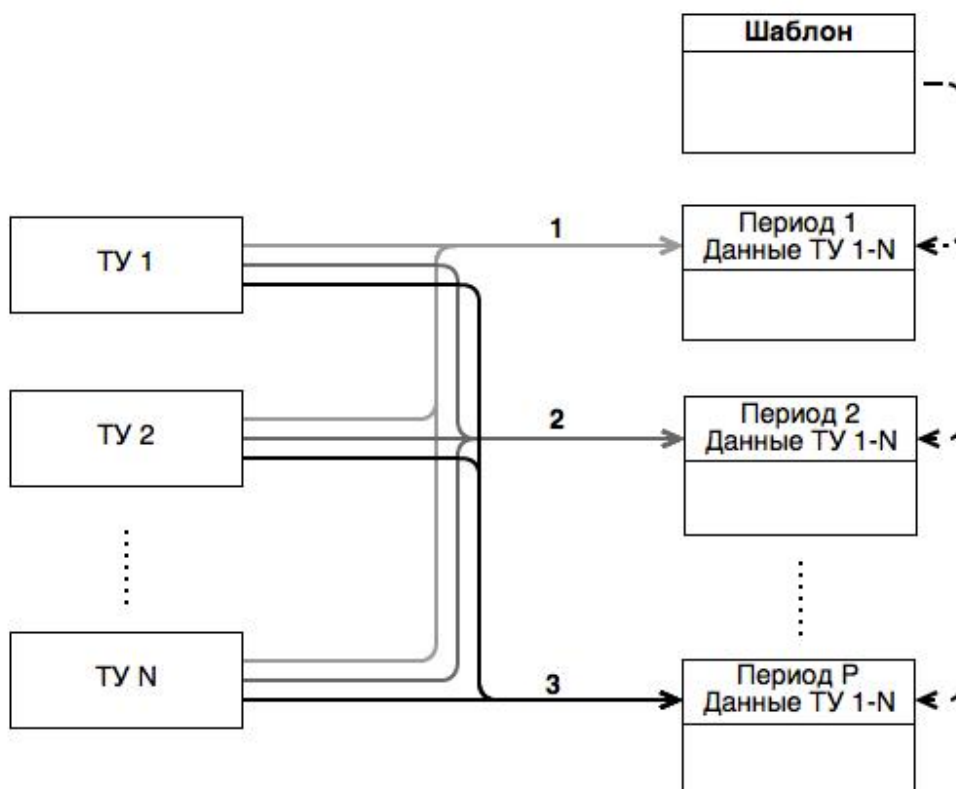


Рисунок 4. Распределение данных поступающих от различных точек учета по периодам

Где линия под номером 1 показывает поток данных первого периода, когда данные ТУ 1-N (всех точек учета) попадают в первую таблицу, созданную по общему шаблону таблиц хранения данных. По окончании первого периода создается вторая таблица (по тому же шаблону таблиц) и поток данных точек учета ТУ 1-N переправляется в нее. Линия 2 – поток данных второго периода разбиения. Аналогично для последующих периодов.

Очевидно, что при увеличении количества узлов учета оптимальный период разбиения пропорционально сокращается. Использование малых временных периодов не всегда удобно, но при отсутствии возможности наращивания памяти и невозможности увеличить периодичность опроса (иначе интенсивность λ), оно может оказаться единственным выходом.

В случае масштабирования данной системы также как и в первом случае возрастет интенсивность поступления записей. Аналогично первому рассмотренному случаю решение проблемы путем увеличения памяти, является также не достаточной удобным и оперативным, но применив второй метод, использующий перерасчет предельно допустимой интенсивности поступления записей, можно получить достаточно масштабируемую систему. В этом случае после перерасчета допустимой суммарной интенсивности, элемент системы отвечающий за распределение нагрузки также ограничивает поступления данных в базу отсекая часть поступающих данных, при этом гарантируя сохранения производительности системы в целом. Но система в целом не требует как вмешательства извне (нет необходимости добавления и настройки оборудования), так и внесения изменений в структуру базы данных. Что в свою очередь положительно сказывается и на надежности системы в целом.

Заключение

В результате исследования были рассмотрены два варианта решения проблемы масштабирования систем хранения хронологических данных электроэнергетических систем. Основываясь на разделении сохраняемой информации по ТУ и разделении по времени поступления (периодам), генерирующим информацию. Были приведены формулы для расчета предельно допустимой интенсивности поступления данных в систему, при которой гарантируется их сохранение в системе; выведено заключение о трудозатратах организации сбора данных электроэнергетических систем.

Литература

1. Дж. Уорсли, Дж. Дрейк, PostgreSQL. Для профессионалов. - Питер, 2003. - 496 стр.
2. www.postgresql.org [электронный ресурс]
3. Simon Riggs, Hannu Krosing, PostgreSQL 9 Admin Cookbook. - PacktPublishing, 2010. - 360 стр.

4. Фишер А.В. Организация хранения хронологических данных в базах данных систем мониторинга и прогнозирования / А.В. Фишер, Р.А. Дьяченко, И.С. Лоба // Научный журнал КубГАУ [Электронный ресурс]. - Краснодар: КубГАУ, 2012. - № 79(05). - Шифр Информрегистра: 04201200012/0376. – Режим доступа: <http://ej.kubagro.ru/get.asp?id=2223>