

АРХИТЕКТУРНЫЕ ПОСТРОЕНИЯ МНОГОМАШИНОЙ ПАРАЛЛЕЛЬНОЙ ОБРАБОТКИ ДАННЫХ В СИСТЕМАХ ЭКОЛОГИЧЕСКОГО МОНИТОРИНГА

Ю.С. Легович, Н.Г. Журавлева, Д.Ю. Максимов, А.В. Сталянский
Институт проблем управления им. В.А. Трапезникова РАН
Россия, 117997, Москва, ул. Профсоюзная, 65
E-mail: legov@ipu.ru, zhurav@ipu.ru, phoenixjhanjaa@yandex.ru

Рассмотрены принципы организации процесса обработки данных в автоматизированных системах экологического мониторинга. Показана возможность распараллелить этот процесс за счет новой организации существующих технических средств и программного обеспечения. Обоснована целесообразность перевода систем экологического мониторинга на массивно-параллельную архитектуру. Предложенная структура позволяет существенно улучшить технико-экономические показатели системы.

ARCHITECTURAL CONSTRUCTIONS IT IS A LOT OF COMPUTER OF PARALLEL DATA PROCESSING IN ECOLOGICAL MONITORING SYSTEMS / Y.S. Legovich, N.G. Zhuravlyov, D.Y. Maxim, A.S. Kudryashov (Institute of Control Sciences RAS, Profsoyuznaya str., 65, Moscow, 117997, Russia). Principles of the organisation of process of data processing in the automated systems of ecological monitoring are considered. Possibility to organise parallel data processing at the expense of the new organisation of existing means and the software is shown. The expediency of transfer of systems of ecological monitoring on massivno-parallel architecture is proved. The offered structure allows to improve technical and economic indicators of system essentially.

Введение

В соответствии с Федеральным законом "Об охране окружающей среды", каждый гражданин имеет право на благоприятную окружающую среду, на ее защиту от негативного воздействия, вызванного хозяйственной и иной деятельностью, чрезвычайными ситуациями природного и техногенного характера, на достоверную информацию о состоянии окружающей среды и на возмещение вреда окружающей среде. Экологический мониторинг - это комплексная система наблюдений за состоянием окружающей среды, оценки и прогноза изменений этого состояния под воздействием природных и антропогенных факторов. Информационной базой данного закона являются системы экологического мониторинга, которые обеспечивают органы государственной власти субъектов Российской Федерации достоверной информацией о видах негативного воздействия на окружающую среду, основными из которых являются:

- выбросы в атмосферный воздух загрязняющих или иных веществ;
- сбросы загрязняющих или иных веществ и микроорганизмов в поверхностные водные и подземные водные объекты и на водосборные площади;
- загрязнение недр, почв;
- размещение отходов производства и потребления;
- загрязнение окружающей среды шумом, теплом, электромагнитными, ионизирующими и другими видами физических воздействий.

Информация о состоянии окружающей среды и ее изменениях, полученная при осуществлении мониторинга окружающей среды, используется органами государственной власти Российской Федерации, органами государственной власти субъектов Российской Федерации, органами местного самоуправления для разработки прогнозов социально-экономического развития и принятия соответствующих решений, разработки федеральных программ в области экологического развития Российской Федерации, целевых программ в области охраны окружающей среды субъектов Российской Федерации и мероприятий по охране окружающей среды. К этому следует добавить, что в соответствии с Федеральным законом негативное воздействие на окружающую среду является платным. Всё сказанное подтверждает актуальность научно-технического направления – разработка и проектирование автоматизированных систем экологического мониторинга.

Особое значение проблема проектирования систем экологического мониторинга приобретает на потенциально опасных объектах, таких как химические, нефтяные, объекты уничтожения химического оружия. Решение задачи снижения экологической напряженности на таких объектах возможно путем создания автоматизированных систем мониторинга окружающей среды, а полученные с помощью этих систем результаты в комплексе с технологическими мероприятиями по повышению экологичности производства позволят эффективно управлять качеством окружающей среды.

Все существующие в нашей стране автоматизированные системы мониторинга действуют на единых принципах, позволяющих провести комплексный анализ и сделать обобщения по данным полученных измерений, дать оценку и прогноз происходящих изменений в экосистемах.

Рассмотрим структуры отечественных автоматизированных систем экологического мониторинга наиболее широко распространенных в настоящее время.

1. Типовая структура автоматизированной системы экологического мониторинга.

Структура типовой Системы Экологического Мониторинга (СЭМ) включает набор из n пунктов сбора экологических данных и центра хранения, обработки и анализа информации – информационного аналитического центра (ИАЦ). Функционально ИАЦ выполняет централизованную обработку информации, проводимую на одном выделенном узле вычислительной сети, вне зависимости от местоположения исходных данных.

Система экологического мониторинга представляет собой двухуровневую распределенную информационную систему, включающую нижний уровень сбора данных и верхний уровень обработки и анализа данных. Нижний уровень включает стационарные и передвижные станции контроля, результаты измерений с которых передаются на верхний уровень - в центр обработки информации по каналам беспроводной связи. Структура системы экологического мониторинга представлена на рис.1.



Сокращения рис.1:

АСПК – автоматизированный стационарный пост контроля;

АППК - автоматизированный передвижной пост контроля;

ХАЛ - химико-аналитическая лаборатория.

Информационный аналитический центр представляет собой подсистему обработки и анализа данных, АСПК, АППК, ХАЛ - подсистему сбора данных. Основной функцией подсистемы сбора данных является сбор данных с измерительных приборов, технологических контроллеров и датчиков, хранение данных в локальной базе данных и выполнение транзакций, связанных с передачей данных на верхний уровень (сервер ИАЦ). Если сетевое соединение между системой сбора данных и сервером ИАЦ потеряно, что на практике встречается довольно часто, или заблокировано, транзакции данных запоминаются в системе сбора данных и восстанавливаются (путем повтора транзакций) на сервере ИАЦ сразу после восстановления соединения.

Все посты контроля (ПК) подсистемы сбора данных информационно объединены с ИАЦ посредством беспроводной локальной сети по топологии типа звезда. Стационарный пост - это специально оборудованное вандалоустойчивое помещение, снабженное системами поддержания микроклимата, пожаротушения, автономного энергопитания, контроля доступа и связи. Внутри стационарного поста размещается аппаратура, необходимая для регистрации концентраций загрязняющих веществ, а также метеорологических параметров. Место для установки стационарного поста выбирается с учетом метеорологических условий и уровней загрязнения атмосферного воздуха. Кроме того, стационарные автоматические посты контроля устанавливаются в жилых зонах населенных пунктов и санитарно-защитных зонах промышленных предприятий. Пост контроля обеспечивает непрерывное автоматическое измерение большого количества параметров: концентраций окиси углерода, окислов азота, диоксида серы, сероводорода, аммиака, диоксида углерода, озона, пыли, формальдегида, а также температуры и относительной влажности атмосферного воздуха, атмосферного давления, скорости и направления ветра, радиационного фона.

Численность населения, тыс. чел.	Количество постов
<50	1
50-100	2
100-200	3
200-500	3-5
500-1000	5-10
1000-2000	10-15
>2000	15-20

При прочих равных условиях количество пунктов определяется исходя из численности населения в соответствии с таблицей 1.

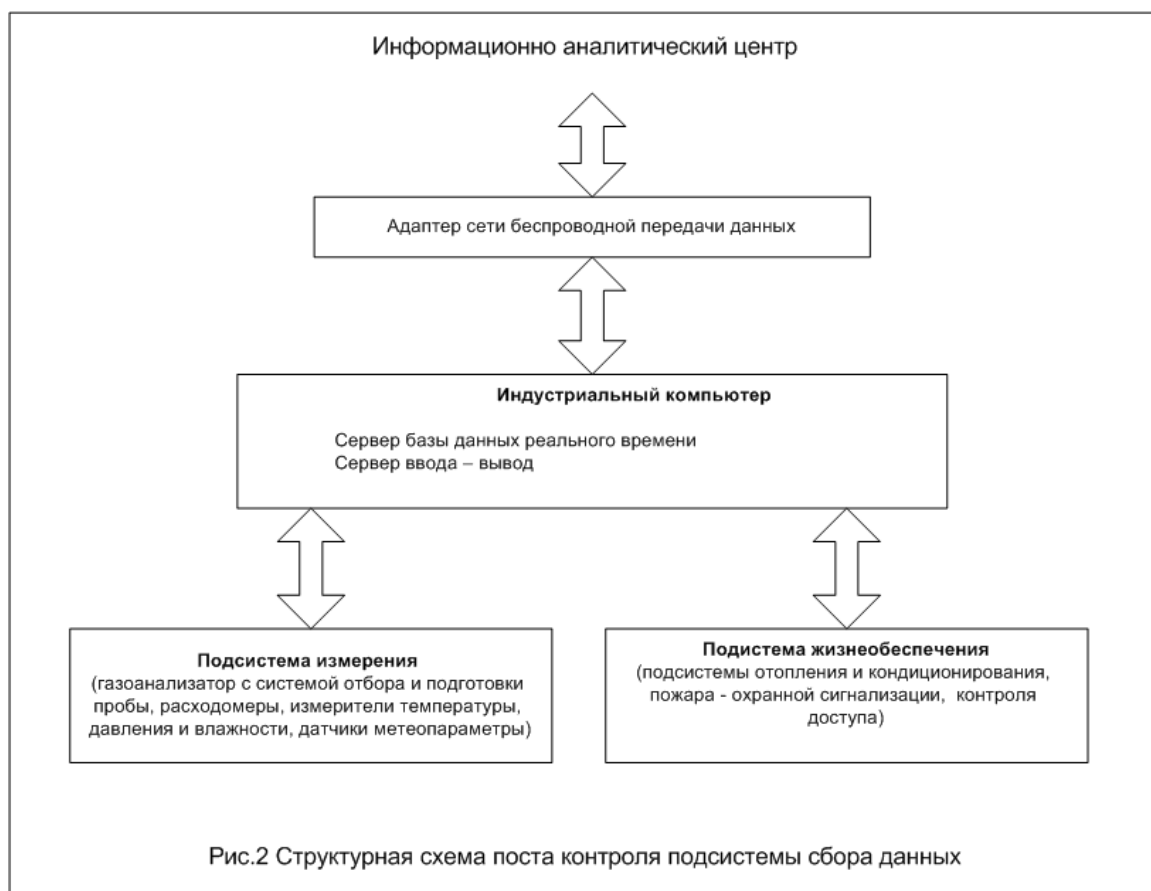
Для населенных пунктов со сложным рельефом и большим числом источников рекомендуется устанавливать один пост на каждые (5-10) км². Отечественной промышленностью выпускаются стандартные павильоны-посты наблюдений или комплектные лаборатории типа экологический стационарный пост, а также разнообразные типы передвижных экологических лабораторий. Эти посты оснащены набором различных приборов и оборудования для отбора проб воздуха, проведения метеорологических измерений: скорости и направления ветра, температуры, влажности.

АППК - автоматизированный передвижной пост контроля может выполнять функции маршрутного или подфакельного поста. Маршрутный пост наблюдений предназначен для регулярного отбора проб воздуха в фиксированной точке местности, расположенной на определенном маршруте. Такой пост имеет производительность около 5000 отборов проб в год, а в день производится отбор 8 - 10 проб воздуха. Подфакельный пост

предназначен для отбора проб под дымовым (газовым) факелом с целью выявления зоны влияния данного источника. Отбор проб воздуха производится последовательно по направлению ветра на расстояниях (0,2 - 0,5); 1, 2, 3, 4, 6, 8, 10, 15 и 20 км от стационарного источника выброса, а также с наветренной стороны источника.

Функционально любой пост контроля (рис.2) включает следующие подсистемы:

- подсистему измерения экологических и метеорологических параметров;
- подсистему жизнеобеспечения – отопления и кондиционирования, контроля возникновения пожара – охранной сигнализации;
- подсистему обработки, регистрации и архивирования информации, поступающей от измерительных приборов и средств жизнеобеспечения, которая состоит из промышленного компьютера с комплектом периферийного оборудования и программного обеспечения.



Программное обеспечение включает базу данных реального времени, сервер ввода-вывода, управляющий вводом данных, поступающих от подсистем измерения, и передачей данных по каналам связи с ИАЦ. Оборудование постов такого типа является весьма дорогостоящим, требует высоких текущих затрат на запчасти и обслуживание, и недостаточно надёжна. Это связано с тем, что датчики анализатора быстро загрязняются и выходят из строя под воздействием содержащихся в выбросах агрессивных компонентов, и поэтому требуют частых проверок и многочасовых калибровок.

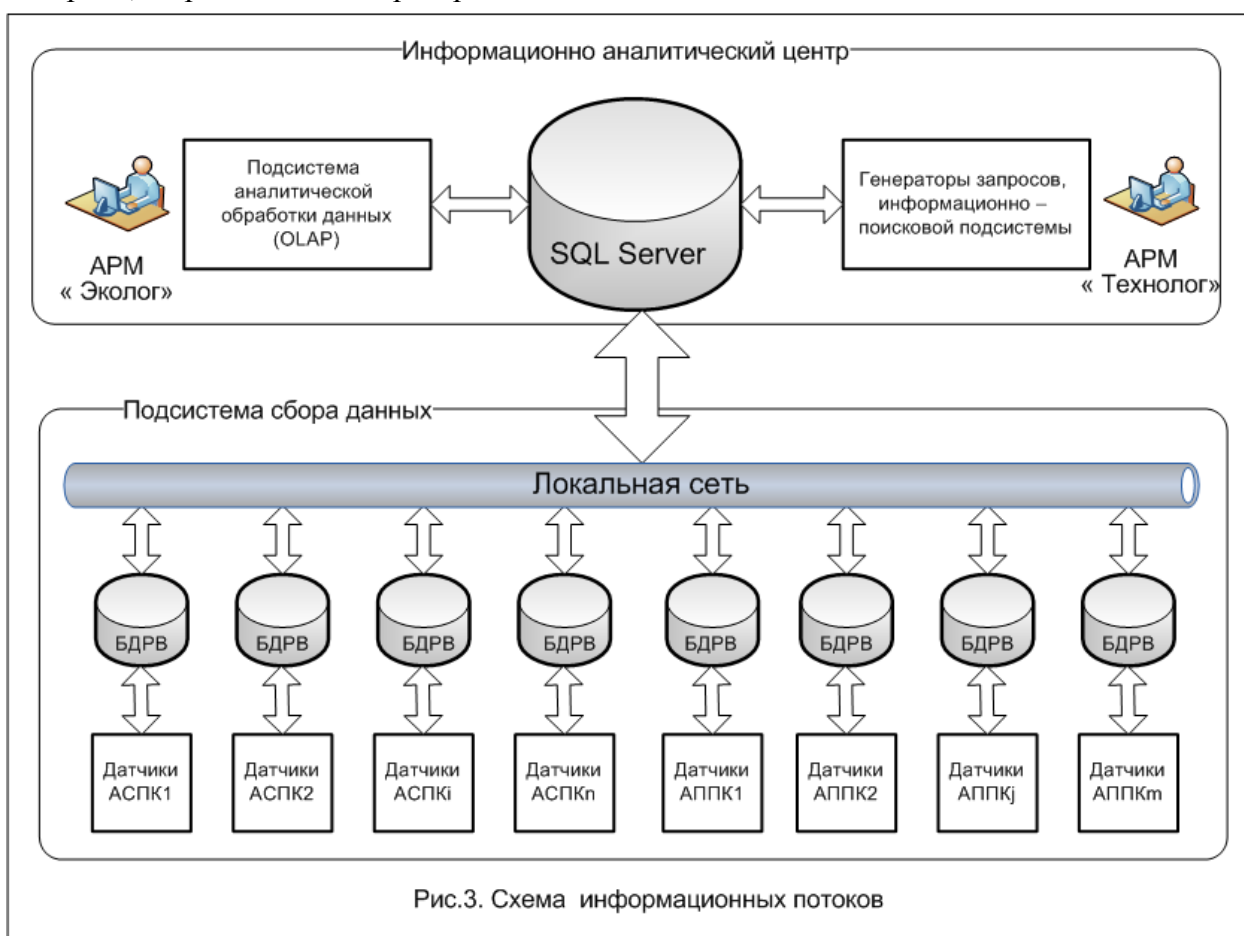
Информационно аналитический центр представляет собой комплекс технических и программных средств, выполняющий следующие функции:

- сбор экологической и технологической информации, поступающей от подсистемы сбора данных;
- обработку, накопление и архивирование данных измерений;
- периодическое формирование информационных сообщений в различные надзорные региональные органы власти;

- информационный поиск и доступ к архивной информации;
- математическое моделирование экологических процессов ближнего переноса и трансформации загрязнений, анализ и прогноз динамики загрязнений;

2. Информационная структура системы мониторинга.

Поток информации — это группа данных, рассматриваемых в процессе ее движения в пространстве и времени в одном направлении. Схема информационных потоков системы экологического мониторинга является однонаправленной - от постов контроля к информационно аналитическому центру. У этих данных есть общий источник и общий приемник. На рис.3 представлена схема информационных потоков системы экологического мониторинга, где источником информации является совокупность постов контроля, а приемником операторы АРМов эколога и технолога.



Информационно управляющая система поста контроля строится на основе промышленного компьютера и выполняет следующие функции:

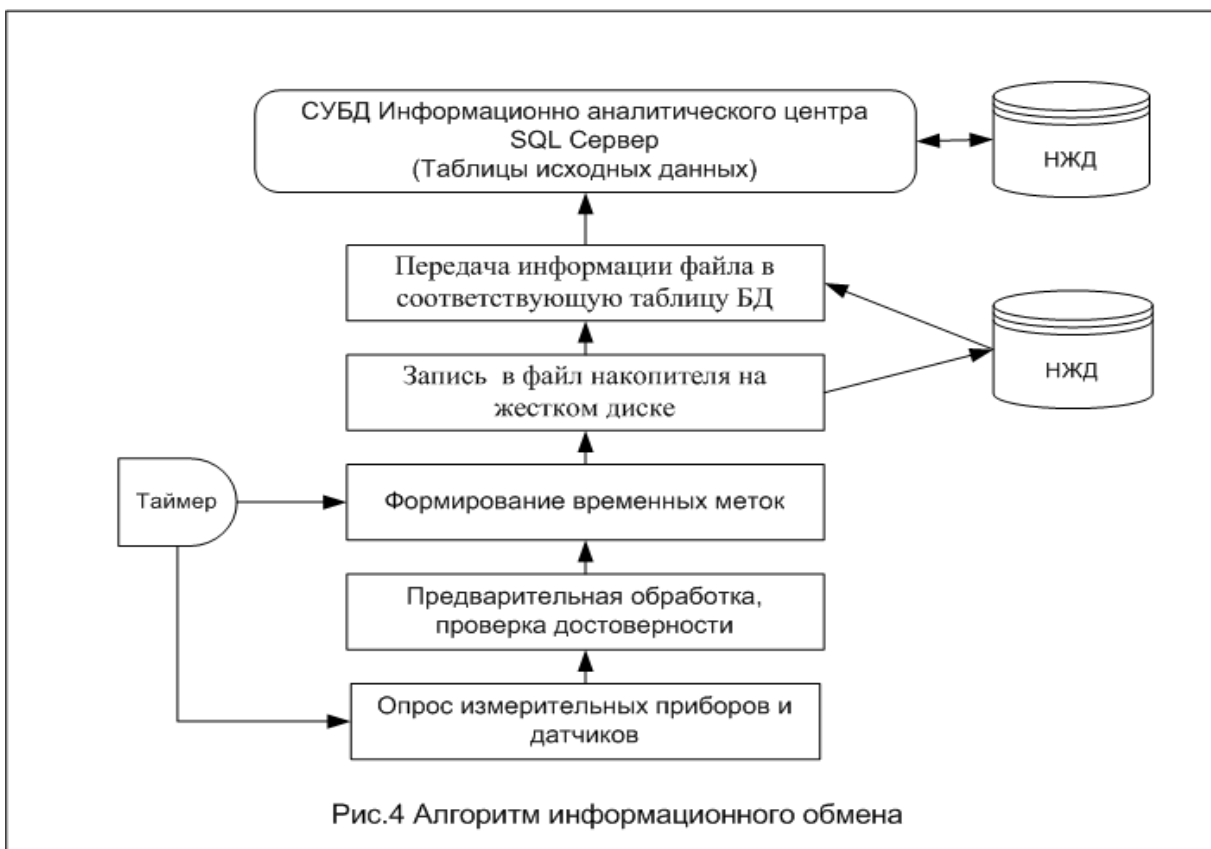
- опрос приборов и датчиков измерительной подсистемы;
- управление режимами измерений;
- выполнение операций по первичной обработке измерительных данных с целью оценки достоверности информации;
- автоматизация проверок и калибровок газоанализаторов;
- управление процессом гарантированной передачи информации в ИАЦ;
- управление функционированием базы данных реального времени;
- управление системой жизнеобеспечения.

На каждом посту контроля в процессе круглосуточной работы периодически с интервалом от 5 до 20 секунд происходит опрос датчиков и измерительных приборов. Данные проходят предварительную обработку и записываются в БДРВ - базу данных реального времени, представляющую собой систему файлов, информация в которых

жестко привязана ко времени поступления данных путем добавления к принятой информации временных меток. Основная специфика БДРВ заключается в ее направленности на высокую скорость накопления данных и уменьшения размеров дискового пространства, необходимого для хранения архива данных. С целью защиты от возможных сбоев в работе беспроводной локальной сети информация записывается в архив на локальном диске. Затем информация переписывается по локальной сети на SQL сервер ИАЦ, где заносится в соответствующую таблицу базы данных.

В случае возникновения временных отказов в работе беспроводной сети передачи данных в информационно аналитический центр сервер ввода-вывода приостанавливает процесс передачи до восстановления работоспособности сети.

Помимо этого непрерывно поступает информация от приборов и датчиков подсистемы жизнеобеспечения – отопления и кондиционирования, пожара – охранной сигнализации. Подсистема жизнеобеспечения предназначена для поддержания внутри поста контроля необходимого температурного режима, отработки аварийных ситуаций (возгорание, вскрытие входной двери, нарушения внешнего охранного контура), передачи видеоизображения системы контроля доступа, управления резервной системой электропитания.



Эта информация также регистрируется в архиве на локальном диске, а события, связанные с нарушением условий нормальной эксплуатации, передаются для обработки на сервер информационно аналитического центра. Алгоритм информационного обмена представлен на рис.4.

Аналитическая обработка данных происходит с использованием специализированных приложений-клиентов, OLAP анализа.

3. Параллельная обработка информации

Как было показано выше, основными качествами архитектуры СЭМ являются ее однородность и высокая степень децентрализации, что позволяет рассматривать (СЭМ) как идеальную многомашинную систему, предназначенную для параллельной обработки данных, которая характеризуется:

- физической удаленностью друг от друга компонентов системы,
- хранения и обработкой данных в местах их создания,
- возможностью использования принципа параллельной обработки данных.

Поскольку с ростом числа компьютеров узким местом информационного обмена становится пропускная способность беспроводной вычислительной сети, то одно из решений данной проблемы заключается в приближении процесса обработки к месту получения данных. Это может быть достигнуто за счет отказа от архитектуры с централизованной обработкой данных и перехода к архитектуре системы с распределенной базой данных (рис.5). С этой целью на рис.5 введены транзакционные системы в базы данных постов контроля, а в информационно аналитический центр - диспетчер параллельных SQL запросов. Переход к такой архитектуре открывает возможность организации процесса параллельной обработки данных в соответствии с таким типом параллелизма как параллелизм объектов или данных. Как известно такой тип параллелизма имеет место тогда, когда по одной и той же программе может обрабатываться некоторая совокупность данных, поступающих в систему одновременно. В данном случае это так и есть, поскольку программы обработки сигналов от датчиков и приборов, установленные на однотипных объектах (посты экологического контроля) одинаковые. Другими словами, все сигналы поступают от одной и той же совокупности параметров, и обрабатываются по одной и той же программе. Данный тип параллелизма используется при построении двух широко распространенных типов многопроцессорных вычислительных систем: массивно-параллельных систем (massively parallel processor or MPP) и кластеров (clusters) [1]. Система с массовым параллелизмом представляет собой многопроцессорную или многомашинную систему с распределенной памятью, в которой с помощью некоторой коммуникационной среды объединяются однородные вычислительные узлы. Каждый из узлов состоит из одного или нескольких процессоров, собственной оперативной и внешней памяти, коммуникационного оборудования, подсистемы ввода/вывода, т.е. обладает всем необходимым для независимого функционирования. При этом на каждом узле может функционировать либо полноценная операционная система (ОС), либо урезанный вариант, поддерживающий только базовые функции ядра, а полноценная ОС работает на специальном управляющем компьютере. Процессоры в таких системах имеют прямой доступ только к своей локальной памяти, а доступ к памяти других узлов реализуется с помощью механизма передачи сообщений. Такая архитектура вычислительной системы устраняет проблему конфликтов при обращении к памяти, что дает возможность практически неограниченного наращивания увеличения числа узлов в системе, увеличивая тем самым ее производительность и обеспечивая высокий уровень масштабируемости. MPP система идеально подходит для случая параллельного выполнения независимых программ, поскольку каждая программа выполняется на своем узле и никаким образом не влияет на выполнение других программ [2].

Другой вариант параллельного выполнения независимых программ состоит в организации кластера. Кластер представляет собой множество компьютеров, объединенных в сеть при помощи специальных аппаратно-программных средств [3]. Кластеры могут быть образованы на базе уже существующих у потребителей отдельных компьютеров, либо же сконструированы из типовых компьютерных элементов, что обычно не требует значительных финансовых затрат. Применение кластеров может также

в некоторой степени снизить проблемы, связанные с разработкой параллельных алгоритмов и программ, поскольку повышение вычислительной мощности отдельных процессоров позволяет строить кластеры из сравнительно небольшого количества (несколько десятков) отдельных компьютеров. Это приводит к тому, что для параллельного выполнения в алгоритмах решения вычислительных задач достаточно выделять только крупные независимые части расчетов, что, в свою очередь, снижает сложность построения параллельных методов вычислений и уменьшает потоки передаваемых данных между компьютерами кластера.

Однако, следующие две особенности кластерных систем затрудняют использование данного типа параллелизма в системах экологического мониторинга. Эти особенности следующие:

1. Все компьютеры в кластере являются равнозначными.
2. Эффективность работы в полной мере зависит от пропускной способности сети, определяющей скорость передачи информации по каналам связи, объединяющим кластеры.

Таким образом, можно сделать вывод, что массивно-параллельные системы (МРР) в большей степени соответствуют представленной на рис.5 структурной схеме, чем кластерные системы. Дополнительным аргументом в пользу выбора МРР архитектуры является возможность реализации на ее основе распределенных вычислений. Данный термин (распределенные вычисления) обычно используют для указания параллельной обработки данных, при которой используется несколько обрабатывающих устройств, достаточно удаленных друг от друга, в которых передача данных по линиям связи приводит к существенным временным задержкам. Как результат, эффективная обработка данных при данном способе организации вычислений возможна только для параллельных алгоритмов с низкой интенсивностью потоков межпроцессорных передач данных.

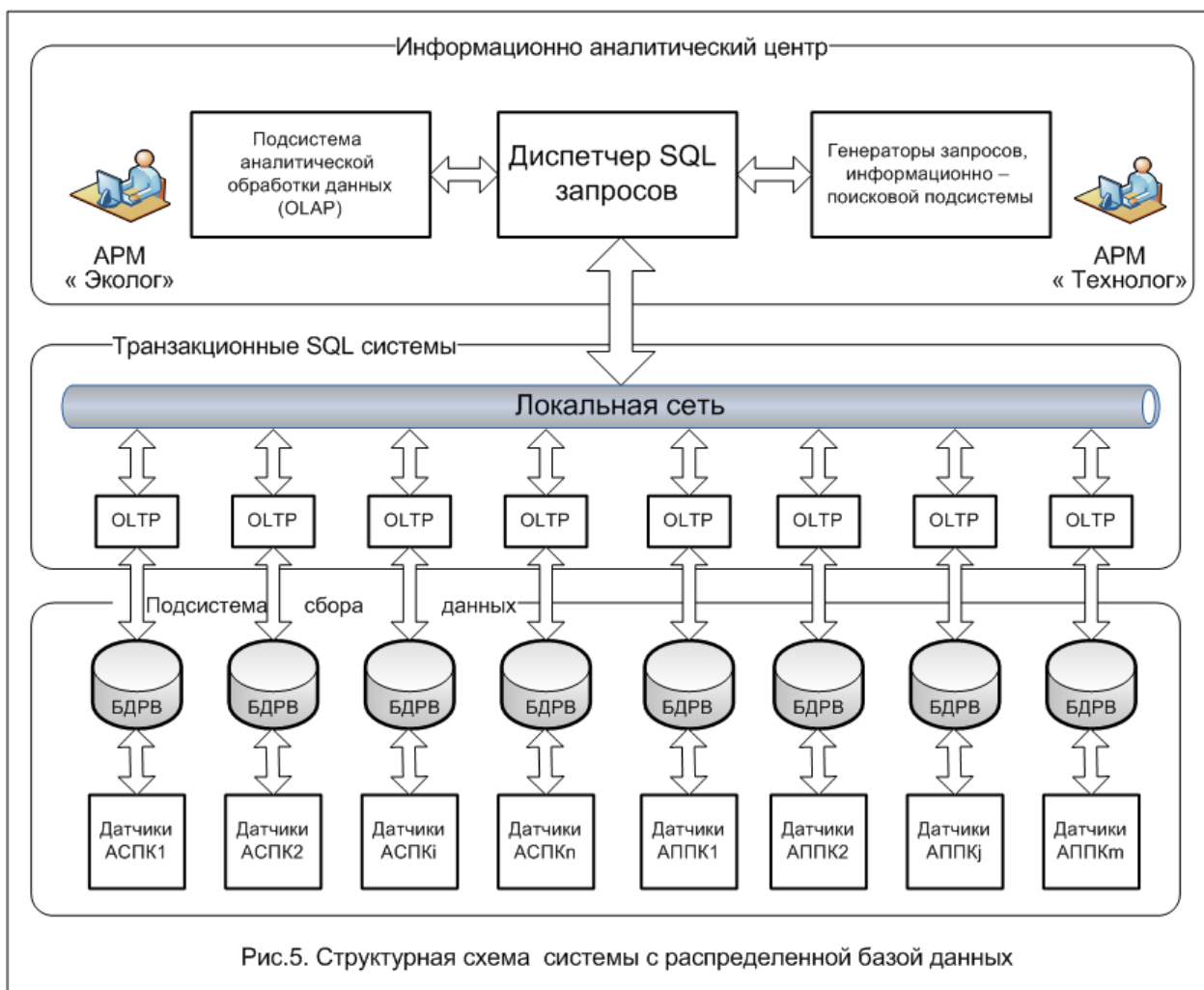
Проведем оценку увеличения производительности при переходе на архитектуру системы экологического мониторинга с параллельной обработкой данных. Для получения предварительной оценки будем использовать известный закон Амдала[4], который Джин Амдал сформулировал еще в 1967 году: «В случае, когда задача разделяется на несколько частей, суммарное время ее выполнения на параллельной системе не может быть меньше времени выполнения самого длинного фрагмента». Если не распараллеливаемая часть кода f может быть равномерно формально распределена по n процессорам, то относительное ускорение S может быть вычислено по следующей формуле:

$$S \leq \frac{1}{f + \frac{1-f}{n}}$$

Следовательно, необходимо найти оценку для f . Исходной информацией может служить работа системы экологического мониторинга традиционной архитектуры с централизованной СУБД. Опыт эксплуатации систем экологического мониторинга объектов уничтожения химического оружия с СУБД на основе Microsoft SQL Server показал, что усредненное количество запросов к базе данных, связанных с обработкой исходных данных принятых от постов контроля составляет 73% от общего количества запросов к СУБД [5].

Количество таких запросов в течение одной рабочей смены (8 часов) составляет 89% от всего количества запросов обработанных СУБД (характеристика относительной интенсивности обращений к СУБД).

Загруженность беспроводной сети со скоростью передачи 54Mbps составила 63%.



В таблице 2 приведены расчеты ускорения работы СУБД в зависимости от доли не распараллеливаемых программ. В таблице: N – количество постов контроля, $K1$ и $K2$ – доля последовательных программ в процентах ($K1 = 27\%$, $K2 = 11\%$).

Данные, приведенные в таблице 2, дают заниженную оценку, поскольку не учитывают характерные для СУБД временные задержки, связанные с работой дисковой памяти. В случае распараллеливания эти задержки уменьшаются обратно пропорционально количеству пунктов контроля, работающих параллельно.

Таблица 2

ПК	Ускорение	
	$K1$	$K2$
N		
5	2.4	3.5
10	2.9	5.0
15	3.1	5.9
20	3.3	6.4
30	3.4	7.1
50	3.5	7.8

Заключение

Рассмотренные принципы организации процесса обработки данных в автоматизированных системах экологического мониторинга позволили доказать возможность распараллелить этот процесс за счет новой организации существующих технических средств и программного обеспечения. Предложенная структура с параллельной обработкой данных позволяет существенно улучшить технико-экономические показатели системы:

- значительно увеличить быстродействие обработки SQL запросов за счет параллелизма;
- обеспечить удобство и простоту масштабирования системы, из-за отсутствия необходимости перепрограммирования;
- снизить общую стоимость технического и программного обеспечения за счет отказа от необходимости использования в информационно аналитическом центре высокопроизводительного сервера баз данных;
- снизить сетевой трафик, что позволяет использовать беспроводные GSM сети или увеличить дальность при использовании WiFi сетей.

Системы экологического мониторинга на основе архитектуры с параллельной обработкой данных обладают значительными технико-экономическими преимуществами по сравнению с традиционной архитектуры. Оценки, проведенные для объектов уничтожения химического оружия, показали, что ожидаемое сокращение экономических затрат от перехода на архитектуру с параллельной обработкой данных составит не менее 35% только за счет снижения технических требований к производительности и объемам памяти сервера баз данных ИАЦ.

Список литературы

1. Корнеев В. Г. Параллельные вычислительные системы. М.: "Нолидж", 1999.
2. Воеводин В.В., Воеводин Вл.В. Параллельные вычисления. -СПб: ВHV, 2002.
3. Baker, M. (2000) *Cluster Computing White Paper*, UK: University of Portsmouth, p. 119.
4. Mark D. Hill and Michael R. Marty. *Amdahl's Law in the Multicore Era*, IEEE Computer, July 2008.
5. Ю.С.Легович, Д.Б.Рождественский. Обработка и анализ данных в системе экологического мониторинга объектов по уничтожению химического оружия. Сборник докладов 8-й Международной конференции "Цифровая обработка сигналов и её применение". –М., 2006, 15-19с.

