

# Автоматизированная система управления данными, максимизирующая скорость доступа к архиву ДЗЗ

В.П. Саворский

*Институт радиотехники и электроники РАН, Фрязинский филиал  
141190 Фрязино, Московской обл., пл. Введенского, 1  
E-mail: [savor@ire.rssi.ru](mailto:savor@ire.rssi.ru)*

Цель работы – повышение эффективности использования архива дистанционного зондирования Земли (ДЗЗ). Проект направлен на решение базовых задач ОКР «Регион-архив», входящей в ФКП, а именно на уменьшение времени доступа к данным, хранящимся в архиве информационной системы (ИС) ДЗЗ. Предлагаемый в работе подход заключается в создании автоматизированных средств переноса (миграции) данных между оперативной частью архива, имеющем средства быстрого онлайн-доступа, и его долговременным отделом. При этом оптимизация процессов миграции данных достигается за счет учета реальной эмпирически устанавливаемой частоты обращения к данным, хранящимся в архиве.

## Введение

Архив ИС ДЗЗ должен обеспечивать:

1. быстрый доступ к актуальным данным,
2. долговременное гарантированное хранение всех включенных в архив наборов данных.

Условия 1 и 2 внутренне противоречат друг другу. Поэтому для разрешения этого противоречия в архиве выделяют 2 раздела:

1. оперативный, обеспечивающий работу с актуальными данными,
2. долговременный, обеспечивающий долговременное хранение данных.

Оперативный раздел архива размещается на средствах, обеспечивающих максимальную для данного архива скорость доступа к данным.

**Основной задачей** работы является разработка автоматизированной системы, управляющей размещением данных в архиве. Актуальные, т.е. поступающие в архив и выдаваемые из архива, данные первоначально размещаются в оперативном разделе архива, а дальнейший их перенос в долговременный архив производится автоматически разрабатываемой в рамках данной работы системой.

## 1. Требования к системе управления данными

Максимизация скорости извлечения данных из архива эквивалентна минимизация времени, требуемой для переноса данных с архивных носителей на средства передачи данных потребителям. Для архива, состоящего из оперативного и долговременного разделов, которые размещены на носителях с существенно различающимися скоростями доступа, она может быть достигнута за счет того, что размещение данных в разделах архива производится с учетом частоты извлечения этих данных по запросам пользователей. При этом максимизация скорости доступа достигается за счет переноса наименее популярных данных на носители с малыми скоростями, которыми оснащен долговременный раздел архива. При этом должны быть учтены следующие факторы:

- 1) наличие в архиве средств с разными скоростями доступа, на пример, массивов дисковых накопителей и высокоплотных магнитных лент,
- 2) размещение всего оперативного раздела архива на дисковых накопителях,
- 3) наряду с размещением в оперативном разделе архива актуальных вновь полученных из системы обработки данных размещение в нем и файлов с данными, имеющими самую высокую частоту пользовательских запросов,
- 4) перенос (миграция) данных из оперативного раздела архива в его долговременный раздел

должны инициироваться автоматически по достижению максимально допустимого уровня заполнения оперативной части архива,

5) миграция данных из оперативного в долговременный раздел архив должна прекратиться автоматически по снижению уровня заполнения оперативного раздела архива до минимально допустимого уровня,

6) данные переносятся в последовательности, обратной популярности данных, оценкой которого является предложенный и обоснованный в работе ранг данных в архиве, определяемый частотой запросов и датами генерации данных и поступления последнего запроса на них.

## 2. Принципы оптимизации размещения данных

Принципы функционирования системы миграции данных из оперативного в долговременный раздел архива ДЗЗ, представлены на рис.1. Они заключаются, прежде всего, в том, что данные переносятся из оперативного раздела в долговременный тогда, когда снижается их актуальность, т.е. когда вероятность поступления заказа на эти данные минимальна по сравнению со всеми данными, хранимыми в оперативном разделе. Обратный перенос производится тогда, когда поступает заказ на данные, которые в момент оформления заказа хранятся в долговременном разделе архива. При этом наряду с переносом данных в оперативный раздел в каталог заносится дата последнего заказа и обновляется значение полного числа заказов на этот набор данных за все время хранения этого набора в архиве. Наряду с датой поступления данных в архив, указанные параметры, как показано далее, необходимы нам для оценок вероятности поступления заказа.

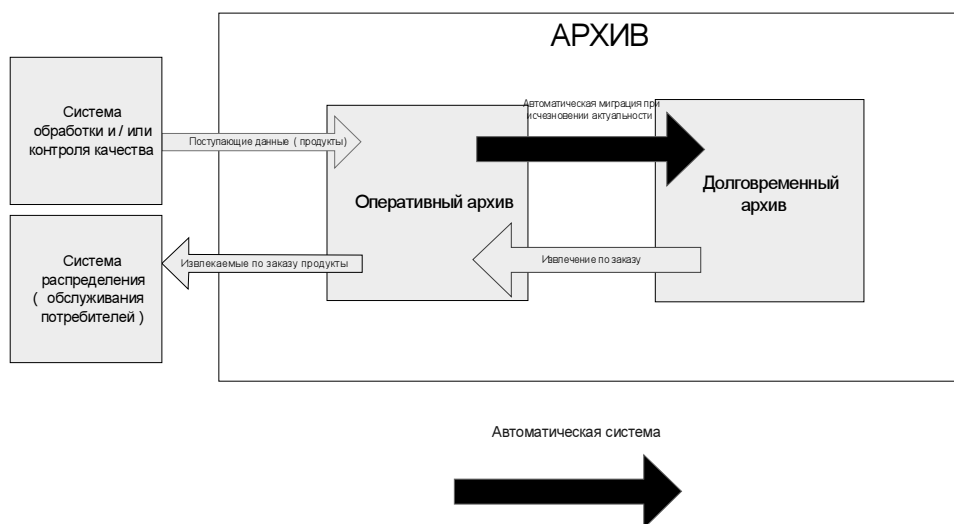


Рис. 1. Функциональная схема работы автоматизированной системы миграции данных в архиве ДЗЗ



Рис. 2. Порядок запуска и останова процедур миграции

Принципы запуска и останова работы системы миграции данных (см. рис.2):

- система миграции данных из оперативного архива в долговременный запускается, когда общий объем хранимых в оперативном архиве данных превышает верхний допустимый уровень,
- система миграции данных из оперативного архива в долговременный прекращает работу, когда общий объем хранимых в оперативном архиве данных становится меньшим, чем нижний допустимый уровень.

## 2. Определение последовательности переноса данных

Поскольку в долговременный раздел архива переносятся не все данные - об этом свидетельствует наличие нижнего допустимого уровня хранения данных в оперативном архиве. Миграция производится в соответствии с рангом популярности данных: очередь на перенос формируется в очередности, обратной популярности данных, которую имеют данные с минимальным рангом. В качестве меры популярности в работе используется эмпирическая оценка вероятности поступления запроса: чем выше вероятность поступления запроса, тем выше ранг данных.

Для оценки вероятности поступления запроса мы используем предположение о том, что запросы формируют поток событий, подчиняющийся статистике простого пуассоновского потока. При таком предположении интенсивность потока (количества событий в единицу времени) запросов на  $i$ -й набор данных определяется следующей эмпирической оценкой:

$$\lambda^i = N^i / (t_N^i - t_0^i),$$

Здесь  $N^i$  - полное количество запросов на  $i$ -й набор данных,  $t_0^i$  - время поступления данных в архив,  $t_N^i$  - время поступления  $N^i$ -го запроса. Вероятность поступления запроса на  $i$ -й набор данных при интенсивности потока заказов  $\lambda^i$  в момент времени  $t > t_N^i$  определяется выражением (вероятность единичного события на интервале  $t - t_N^i > 0$ ) [1]:

$$p^i = 1 - \exp\{-\lambda^i (t - t_N^i)\}.$$

Полученные оценки определяют ранг набора данных. Для множества из  $K$  наборов данных ранжирование сводится к сортировке данных в соответствии с правилом:

$$p^i < p^j, \text{ при } i < j \text{ для любых } i \in \overline{1, K}, j \in \overline{1, K}$$

В результате такой сортировки наборов данных набор с индексом  $i$  является искомым набором данных, который необходимо перенести в долговременный раздел архива. В этом случае минимизируются расчетные затраты на повторное извлечение данных из долговременного раздела архива. При реализации данного алгоритма сортировку имеет смысл проводить при потенциально большом количестве кандидатов на перенос. Если же количество переносимых файлов (наборов данных) невелико, то можно ограничиться простым последовательным переносом файлов, имеющих минимальные вероятности поступления заказа.

## 3. Реализация оптимизированной системы размещения данных

Представленный подход реализован в виде набора процедур Perl, обеспечивающих: 1) контроль ресурсов оперативного раздела архива, 2) расчет ранга популярности данных, 3) автоматический перенос данных, инициируемых процедурами контроля состояния оперативного архива.

## Литература

1. Тихонов В.И. Статистическая радиотехника. 2-е изд., перераб. и доп. – М.: Радио и связь, 1982. – 624с.