

Компоновка БД для R/3 под Oracle

Михаил Прусов

23 ноября 2003 г.

Аннотация

Эта статья является пересказом документа «Database Layout for R/3. Installations under ORACLE» написанного специалистами «Группы передовых технологий» (ATG) фирмы SAP. Оригинальный документ Вы можете найти на странице этой группы по адресу <http://service.sap.com/atg>.

Статья предлагает определенную канву при компоновке данных БД Oracle на жестких дисках. В общем, предложения очень близки подходу S.A.M.E корпорации Oracle.

Статью подготовил Михаил Прусов (<http://mprusov.narod.ru/>). По всем вопросам и предложениям пишите по адресу <mailto:mprusov@yandex.ru>.

Содержание

1	Важность распределения данных	2
2	Предпосылки	2
3	Базовые концепции	2
4	Анализ табличных пространств	3
5	Распределение файлов данных	3
6	Журнальные файлы	4
7	Резюме	4

1 Важность распределения данных

Отображение данных на дисковую подсистему является ключевым фактором производительности информационных систем. Вне зависимости от типа системы, используемой для управления данными, они хранятся на внешнем носителе. На сегодняшний день, в качестве внешнего носителя, обычно используется высокоскоростная дисковая система. С увеличением сложности приложений быстро растет объем обрабатываемых данных, поэтому данные в дисковой системе должны быть в определенном смысле распределены, что позволит сделать доступ к ним максимально быстрым.

Приведенные соображения применимы также и к системе SAP R/3. Эта система может использоваться с различными системами управления баз данных, поэтому необходимо принимать во внимание особенности распределения данных для каждой конкретной СУБД. Эта статья рассматривает систему SAP R/3, работающую с БД Oracle.

2 Предпосылки

Стандартная система SAP R/3, после установки, имеет некоторое количество табличных пространств, распределение таблиц по этим табличным пространствам зависит от многих параметров. Изначально, данные разделены в соответствии с их типом так, основные данные и

транзакционные данные хранятся в разных табличных пространствах. При установке системы SAP R/3 не всегда возможно предсказать размер отдельных таблиц, поэтому трудно сделать точное распределение данных в момент инсталляции. Только эксплуатация может показать какие таблицы станут проблемными с точки зрения размера или загрузки системы ввода/вывода. Обычно, такие таблицы переносятся в выделенные табличные пространства. Не систематизированная деятельность администратора может привести в большому числу табличных пространств, случайным образом расположенных по всей дисковой подсистеме. Подход, описанный здесь, направлен на решение подобной проблемы. Целью является оптимизация производительности и минимизация затрат на администрирование.

Компоновка базы данных, описанная здесь, была успешно апробирована на нескольких проектах. Она решила множество хорошо известных проблем производительности, которые возникают при несбалансированном распределении нагрузки на ввода/вывода в базе данных.

3 Базовые концепции

Для максимального использования всех функций БД (таких как асинхронный ввод/вывод, который не реализован на уровне файловых систем в большинстве операционных систем) в большой системе, необходимо использовать линейные устройства (raw devices) для всех объектов базы данных. За одним исключением (оперативные журнальные файлы), линейные устройства должны располагаться на дисковых томах с чередованием, или, если аппаратное чередование не поддерживается, чередование не должно использоваться вовсе. Размер разделов с линейными устройствами зависит от комбинации базы данных и операционной системы.

В дополнение к стандартным табличным пространствам, созданным при инсталляции, необходимо создать два новых (PSAPCUSTD и PSAPCUSTI). Эти табличные пространства должны использоваться для хранения таблиц, характеризующихся большим размером и/или высокой активностью ввода/вывода, и могут достигать нескольких сотен гигабайт в размере. Пре-

имущество такой стратегии заключается в том, что она дает нам возможность легко распределить данные между большим числом дисков.

В общих чертах, все современные подсистемы ввода/вывода имеют одинаковую базовую структуру. Дисковый контроллер обслуживает массив физических дисков. Поддерживается ли чередование на аппаратном уровне, зависит от особенностей конкретного контроллера. Практика показывает, что использование чередования на программном уровне негативно влияет на производительность системы. Если аппаратное чередование не поддерживается, не используйте чередование. Политика выделения пространства на уровне базы данных Oracle компенсирует отсутствие чередования (см. рисунок 2).

В статье используется термин «раздел диска». Если чередование доступно, то «раздел диска» — это дисковый массив с чередованием; если чередование не используется, то «раздел диска» — это раздел физического диска.

Для реализации стратегии, описанной здесь, Вы должны распределить табличные пространства PSAPCUSTD и PSAPCUSTI между максимальным числом дисков. Равным образом должны быть распределены табличные пространства с сегментами отката и временными сегментами. Отдельного рассмотрения требуют оперативные журнальные файлы.

4 Анализ табличных пространств

Если проанализировать БД системы SAP R/3, то можно увидеть, что около 98% таблиц системы практически не растут и не имеют высокого уровня загрузки ввода/вывода. Только около 1% таблиц характеризуются высоким темпом роста и только около 1% таблиц имеют высокую степень загрузки ввода/вывода. Высоким уровнем загрузки ввода/вывода характеризуются табличное пространство с временными сегментами (PSAPTEMP) и табличное пространство с сегментами отката (PSAPROLL). Список 1 дает представление о характеристиках загрузки ввода/вывода для табличных пространств системы SAP R/3.

Все табличные пространства, определенные

как имеющие большое число операций чтения или записи, из приведенного списка, должны быть проанализированы с целью выявления объектов с высоким уровнем активности ввода/вывода. Эти объекты должны быть перенесены в новые табличные пространства PSAPCUSTD и PSAPCUSTI. Выполнение процедуры позволит сделать все обычные стандартные табличные пространства незначительными с точки зрения производительности. Также самое необходимо проделать с быстрорастущими объектами, — их необходимо выявить и перенести в новые табличные пространства.

Итогом деятельности по переносу объектов БД в новые табличные пространства должен быть новый список 2 табличных пространств.

5 Распределение файлов данных

Табличные пространства, которые хранят объекты с большой интенсивностью ввода/вывода или высокой скоростью роста должны быть равномерно распределены между максимально возможным числом дисков. Для достижения равномерного распределения необходимо определить единый размер файлов данных для каждого табличного пространства. Файлы данных помещаются на все дисковые разделы.

Рисунок 1 показывает размещение файлов данных табличного пространства на дисковых массивах с аппаратной реализацией чередования. На рисунке 2 показано размещение файлов данных в случае, когда отсутствует поддержка аппаратного чередования.

Способ размещения файлов данных, при отсутствии аппаратного чередования, основан на особенностях стратегии выделения новых экстенгов объекта сервером Oracle. В базе данных каждый новый экстенг выделяется в новом файле данных по принципу «карусели». Желательно, чтобы число экстенгов каждого объекта было не меньше, чем файлов данных. В идеале, — число экстенгов должно быть кратным числу файлов данных.

6 Журнальные файлы

В виду того, что каждая фиксация транзакции приводит к записи в оперативный журнальный файл сервера Oracle, размещение оперативных журнальных файлов оказывает сильнейшее влияние на скорость выполнения транзакций. В целях надежности, оперативные журнальные файлы дублируются, первая копия называется `origlog`, вторая — `mirrlog`. При проектировании компоновки файлов убедитесь, что копии журнальных файлов расположены на разных дисках. Часто оперативные журнальные файлы из одной группы (копии) располагают на одном диске исходя из предположения, что они не влияют друг на друга. При этом забывают, что такое расположение ведет к конкуренции за диск между процессом LGWR и ARCH сервера Oracle: в то время, как процесс LGWR пишет в журнальный файл `mirrlogA`, процесс ARCH может выполнять архивирование журнального файла `mirrlogB`.

Таким образом, желательно чтобы файлы `origlogA`, `origlogB`, `mirrlogA` и `mirrlogB` располагались на разных дисках. В виде альтернативы, Вы можете поместить `mirrlogA` и `mirrlogB` на один диск. В штатном режиме такая схема не хуже, чем первая. Однако, при отказе одного из дисков с `origlogA` или `origlogB` процессы LGWR и ARCH начнут конкурировать за устройство, с `mirrlogA` и `mirrlogB`.

Опыт показывает, что размер журнальных файлов должен быть примерно равен 128MB, число оперативных журнальных файлов — в зависимости от нагрузки на систему. Если используется, рекомендуемая здесь, схема с выделенными дисками, то число оперативных журнальных файлов может быть достаточно большим.

Файловую систему с архивными журнальными файлами также рекомендуется хранить на выделенных дисках. В системах с очень высокой загрузкой узким местом может оказаться не только процесс LGWR, но и процесс ARCH. Следует обратить внимание на влияние конкуренции за ввод/вывод в файловой системе с архивными журнальными файлами между процессом ARCH и процессом, выполняющим резервное копирование архивных журнальных файлов, на ленточный накопитель. Конкуренция может стать особенно актуальным при использовании одновременно нескольких современных ленточных нако-

пителей. Для минимизации влияния конкуренции на производительность системы рекомендуется использовать чередование при создании файловой системы для архивных журнальных файлов.

7 Резюме

При проектировании компоновки БД для системы SAP R/3 под БД Oracle используйте следующие рекомендации:

- Поместите журнальные файлы на выделенные диски.
- Распределите данные между всеми доступными дисками и контроллерами.
- Используйте одинаковые значения для параметров хранения задающих размер начального экстенда (`INITIAL`) и последующих экстендов (`NEXT`).
- Установите размеры экстендов для табличных пространств `PSAPCUST*` в диапазоне от 4MB до 128MB, в зависимости от размера объекта БД.
- Снимите ограничение на максимальное число экстендов для объектов в табличных пространствах `PSAPCUST*`.

Пример размещения файлов данных на дисковой системе приведен на рисунке 3.

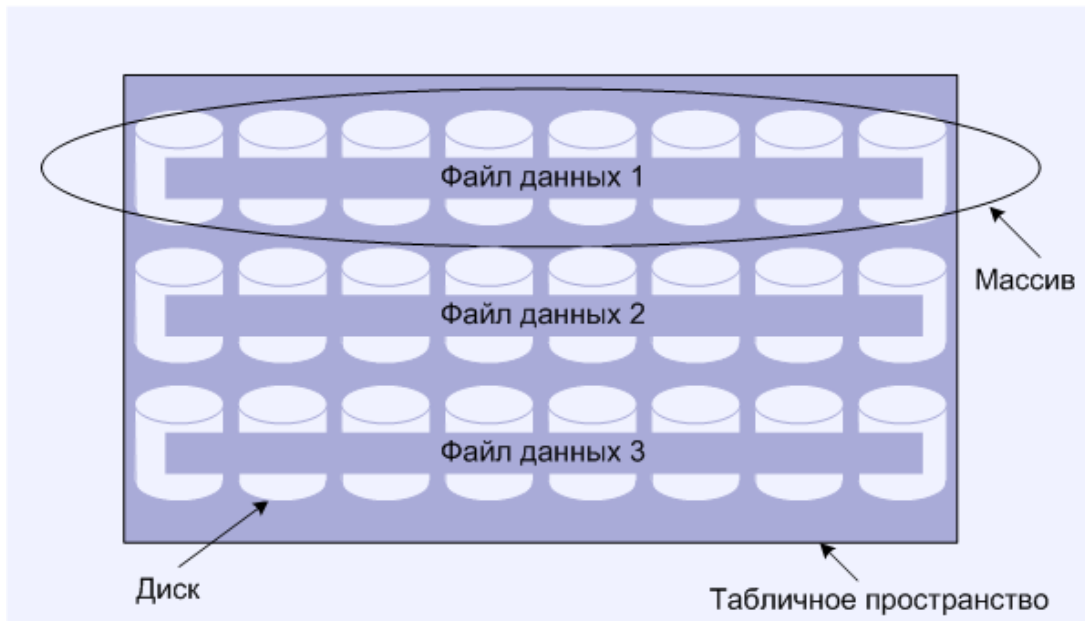


Рис. 1. Размещение файлов данных табличного пространства на дисковых массивах с аппаратной реализацией чередования.

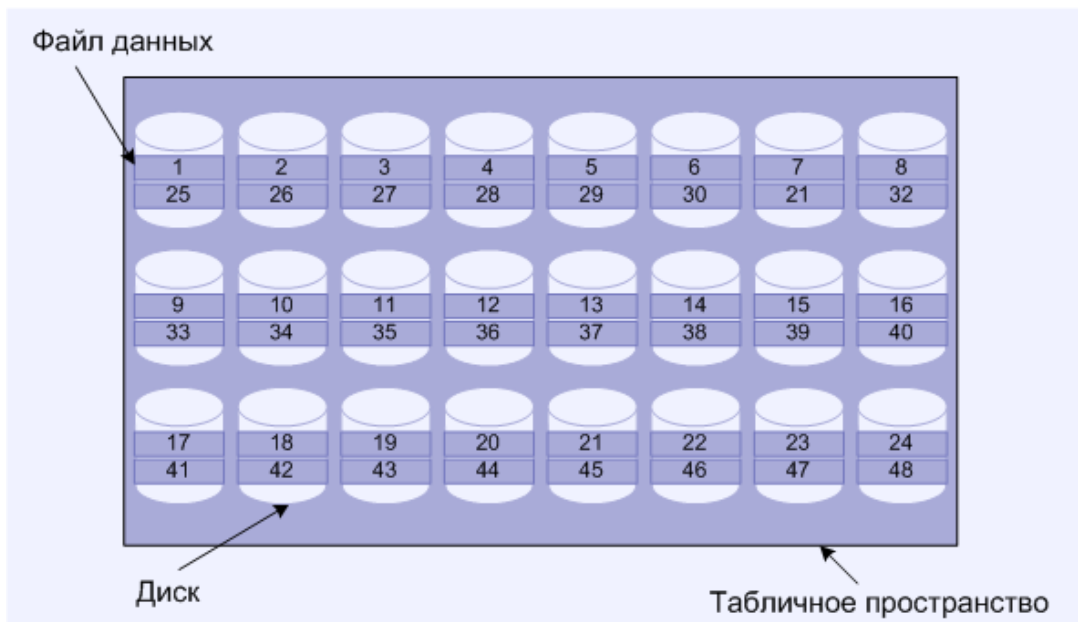


Рис. 2. Размещение файлов данных табличного пространства при отсутствии поддержки аппаратной реализации чередования.

Табличное пространство	Частое чтение	Частая запись	Назначение
PSAPBTABD	Да	Да	Транзакционные данные
PSAPBTABI	Да	Да	Индексы
PSAPCLUD	Да	Да	Транзакционные данные в pool-таблицах
PSAPCLUI	Да	Да	Индексы
PSAPDDICD	Нет	Нет	ABAP-словарь данных
PSAPDDICI	Нет	Нет	Индексы
PSAPDOCUD	Нет	Нет	Документация
PSAPDOCUI	Нет	Нет	Индексы
PSAPELxxxD	Старт	Нет	Скомпилированные программы
PSAPELxxxI	Старт	Нет	Индексы
PSAPESxxxD	Старт	Нет	Исходные тексты программ
PSAPESxxxI	Старт	Нет	Индексы
PSAPLOADD	Старт	Нет	Данные скомпилированных программ
PSAPLOADI	Старт	Нет	Индексы
PSAPPOOLD	Да	Да	Pool-таблицы
PSAPPOOLI	Да	Да	Индексы
PSAPPROTD	Да	Да	
PSAPPROTI	Да	Да	
PSAPROLL	Да	Да	Сегменты отката
PSAPSOURCED	Нет	Нет	Исходные тексты программ
PSAPSOURCEI	Нет	Нет	Индексы
PSAPSTABD	Да	Да	Справочники, прикладные объекты
PSAPSTABI	Да	Да	Индексы
PSAPTEMP	Да	Да	Временные сегменты
PSAPUSER1D	Нет	Нет	таблицы, определенные клиентом
PSAPUSER1I	Нет	Нет	Индексы
SYSTEM	Нет	Нет	Система Oracle

Таблица 1. Стандартные табличные пространства системы SAP R/3, их профиль загрузки ввода/вывода и назначение.

Табличное пространство	Частое чтение	Частая запись	Назначение
PSAPBTABD	Нет	Нет	Транзакционные данные
PSAPBTABI	Нет	Нет	Индексы
PSAPCLUD	Нет	Нет	Транзакционные данные в pool-таблицах
PSAPCLUI	Нет	Нет	Индексы
PSAPDDICD	Нет	Нет	ABAP-словарь данных
PSAPDDICI	Нет	Нет	Индексы
PSAPDOCUD	Нет	Нет	Документация
PSAPDOCUI	Нет	Нет	Индексы
PSAPELxxxD	Старт	Нет	Скомпилированные программы
PSAPELxxxI	Старт	Нет	Индексы
PSAPESxxxD	Старт	Нет	Исходные тексты программ
PSAPESxxxI	Старт	Нет	Индексы
PSAPLOADD	Старт	Нет	Данные скомпилированных программ
PSAPLOADI	Старт	Нет	Индексы
PSAPPOOLD	Нет	Нет	Pool-таблицы
PSAPPOOLI	Нет	Нет	Индексы
PSAPPROTD	Нет	Нет	
PSAPPROTI	Нет	Нет	
PSAPROLL	Да	Да	Сегменты отката
PSAPSOURCED	Нет	Нет	Исходные тексты программ
PSAPSOURCEI	Нет	Нет	Индексы
PSAPSTABD	Нет	Нет	Справочники, прикладные объекты
PSAPSTABI	Нет	Нет	Индексы
PSAPTEMP	Да	Да	Временные сегменты
PSAPUSER1D	Нет	Нет	таблицы, определенные клиентом
PSAPUSER1I	Нет	Нет	Индексы
PSAPCUSTD	Да	Да	Таблицы с высоким уровнем ввода/вывода, быстрорастущие таблицы
PSAPCUSTI	Да	Да	Индексы с высоким уровнем ввода/вывода, быстрорастущие индексы
SYSTEM	Нет	Нет	Система Oracle

Таблица 2. Табличные пространства системы SAP R/3 с новой стратегией размещения объектов БД, их профиль загрузки ввода/вывода и назначение.

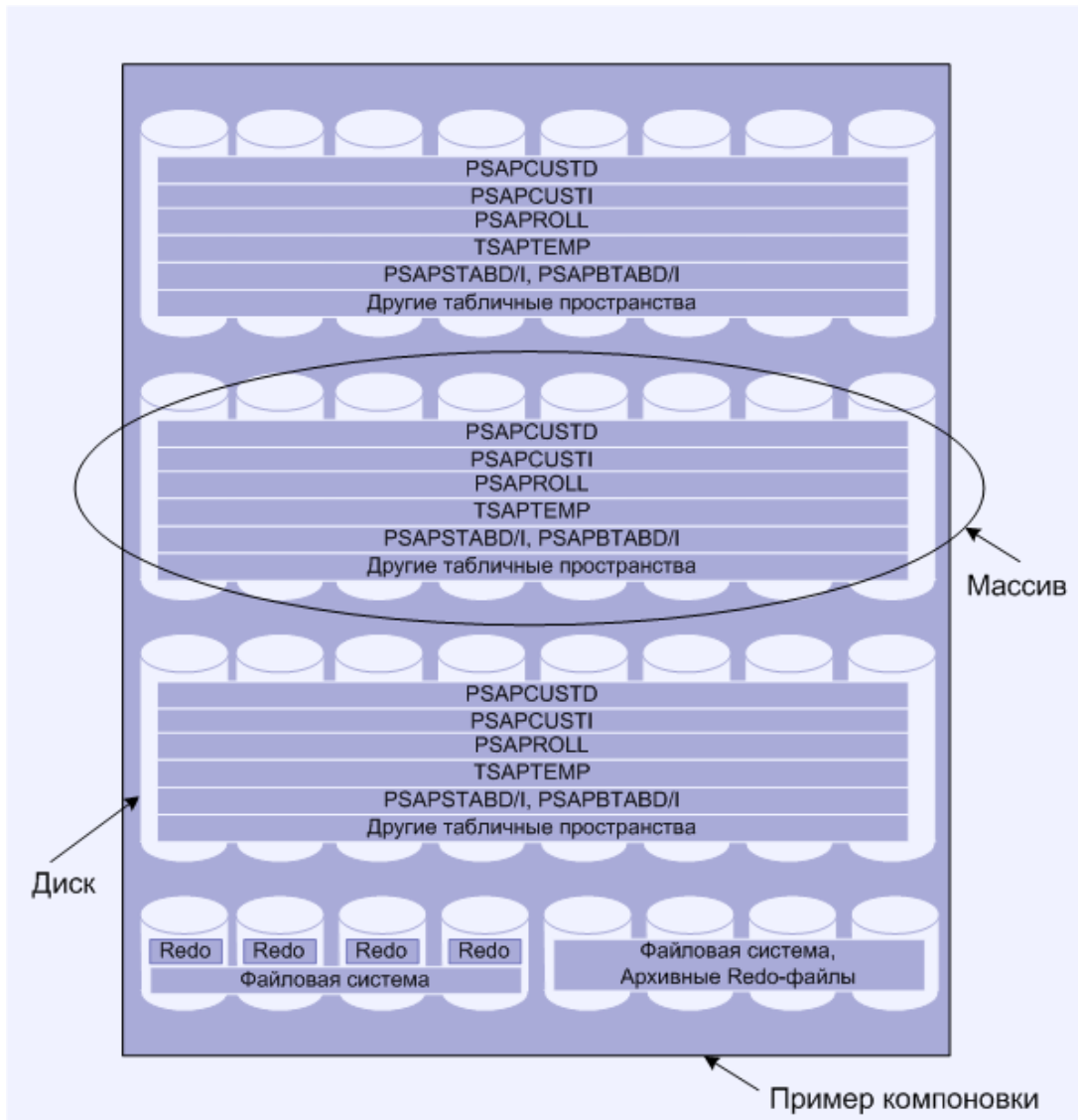


Рис. 3. Пример размещения файлов базы данных.