24 HOURS OF PASS
RUSSIAN 2016

SQL Server Transaction log

Внутреннее устройство и решение проблем

Завадский Андрей

Обо мне

- Архитектор, SQL и .NET разработчик, Аякс-Медиа, Краснодар
- Опыт в IT сфере – 20 лет
  - SQL Server – начиная с версии 7.0 (2001 г.)
  - Разработка: FoxPro, Clipper, Delphi, VB, C#, ASP.NET, MVC, JS, Sharepoint
- Контакты:
  - http://andreyzavadskiy.com
  - https://www.facebook.com/andrey.k.zavadskiy
  - @AndreyZavadskiy
  - https://www.linkedin.com/in/zavadskiy

Содержание

- Логическая и физическая архитектура
- Транзакции и журнал транзакций
- Расширение и усечение журнала транзакций
- Фрагментация журнала транзакций
- Решение проблем
- Отложенные транзакции (Delayed durability)

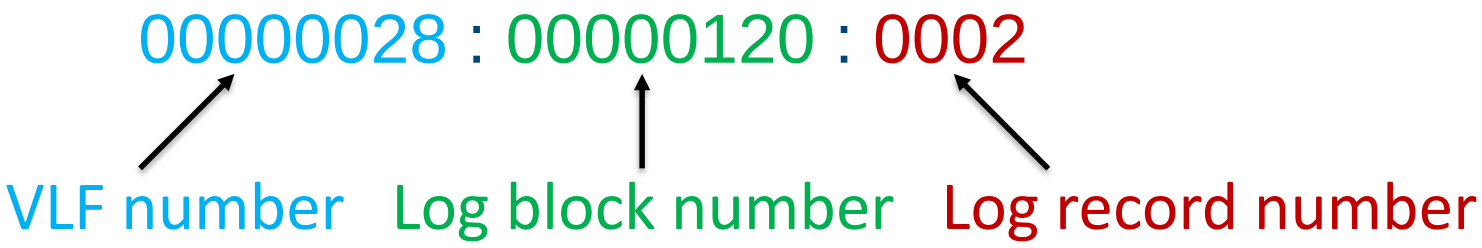
Предназначение журнала транзакций

- Обеспечивает требования ACID (атомарность, согласованность, изолированность и долговечность)
- Восстановление БД в случае разрушения либо при перезапуске SQL сервера
- Восстановление БД, файла, файловой группы, страницы до момента сбоя
- Поддержка транзакционной репликации
- Поддержка решений высокой надежности (HA&DR)

4

Логическая архитектура

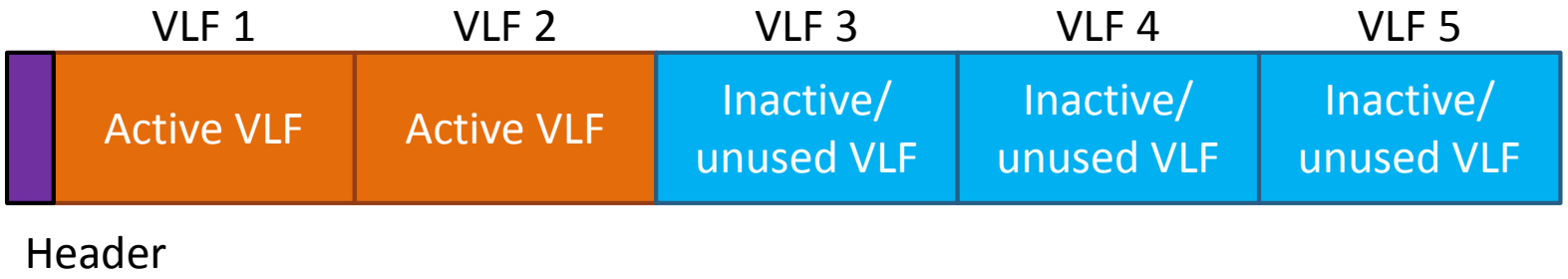
- Представляет собой список записей журнала (log record)
- Каждая запись имеет уникальный номер – Log Sequence Number (LSN)



- Записи журнала хранят информацию о транзакциях, состояниях до и после изменения, размещении (allocation) и др..
- Просмотр доступен через T-SQL функцию fn_dblog()

5

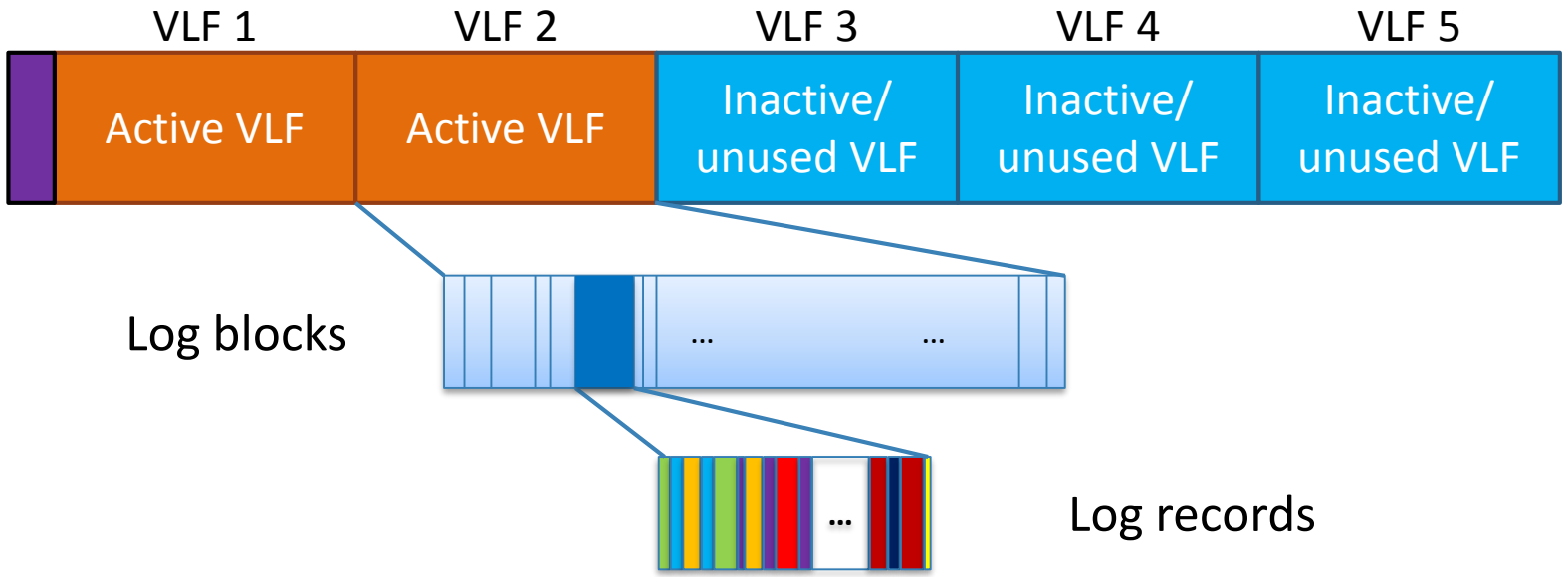
Физическая архитектура (1)



- Заголовок файла 8 Кбайт с метаданными
- Состоит из виртуальных файлов журнала (Virtual Log File, сокр. VLF)
- Идентификатор VLF – логический последовательный номер (FSeqNo)
- Журнал транзакций содержит минимум 2 VLF, размер одного VLF – от 248 Кбайт
- Всегда заполняется нулями при создании

6

Физическая архитектура (2)



- Каждый VLF разбит на блоки журнала (log block)
- Размер блока от 512 байт до 60 Кбайт
- Блок может содержать записи журнала, относящиеся к разным транзакциям

7

Транзакции и журнал транзакций

- Журнал транзакций содержит все модификации, сделанные каждой транзакцией
- Каждая транзакция генерирует несколько записей в журнале
- SQL Server использует журнал с упреждающей записью (write-ahead log)
  - Может быть изменена в режиме отложенных транзакций (Delayed Durability)

8

Демонстрация

- Просмотр записей журнала транзакций

9

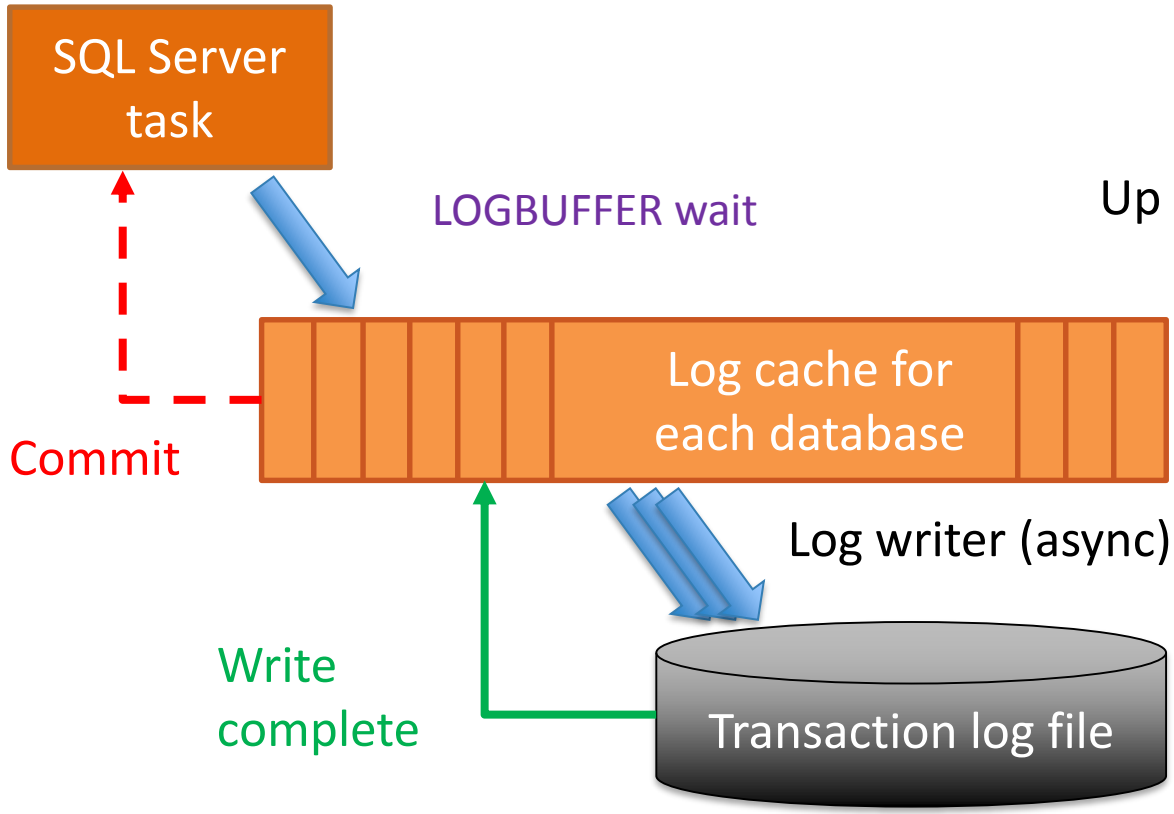
Подтверждение транзакции (commit)

Этапы подтверждения транзакции:

1. Все записи журнала транзакций, включая LOP_COMMIT_XACT, должны быть записаны на диск
2. Ждет подтверждения от сервера, участвующего в синхронном зеркальном отображении БД (synchronous mirror), или сервера AlwaysOn Availability Group
3. Освобождает все блокировки
4. Отправляет подтверждение пользователю

10

Запись транзакций в файл (flush)



Up to 128*60 KB buffers 64 bit
32*60 KB buffers 32 bit

WRITELOG wait

11

Особенности записи транзакций в файл

- Всегда выполняется последовательно
  - Наличие нескольких файлов журнала не дает никаких преимуществ в производительности
- Есть ограничения по пропускной способности записи в файл лога (outstanding I/O)

Версия SQL сервера	Количество потоков ввода-вывода 64 бит	32 бит	Объем данных, Кбайт
до 2005SP1	8	8	480
2005SP1 – 2008R2	32	8	480 (2005) 3840 (2008)
2012 и позже	112	16	3840

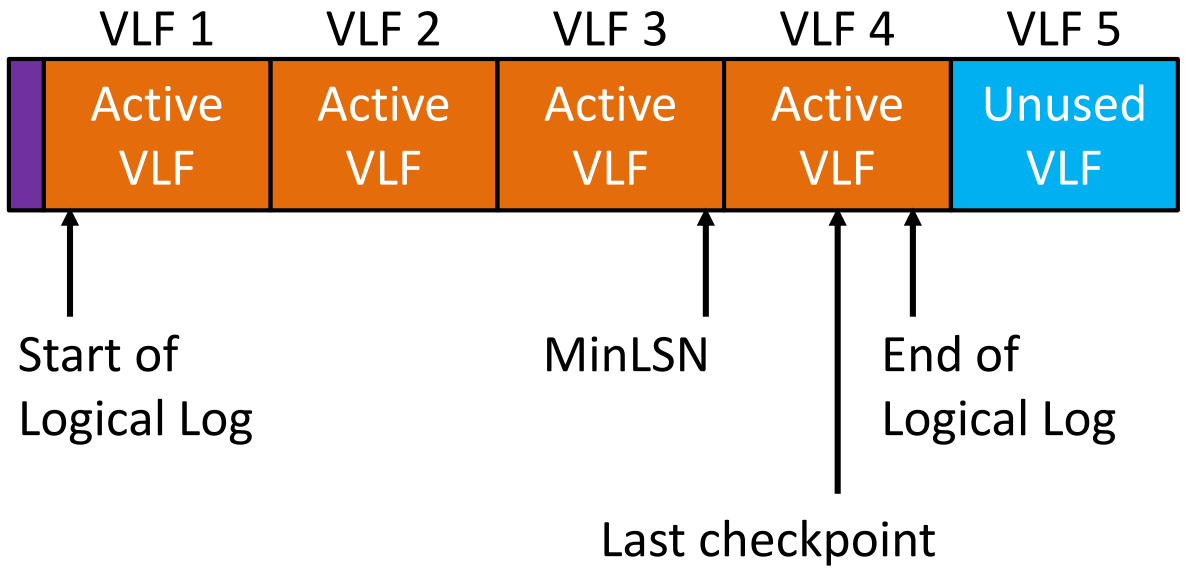
12

Операции с файлами журнала транзакций

Операция	Размер файла	Количество VLF	Примечание
Расширение (Growing)	Увеличивается (в соответствии с параметрами расширения)	Увеличивается (вычисляется по формулам)	
Усечение (truncation)	Не изменяется	Не изменяется	Неактивные VLF помечаются как усеченные
Сжатие (shrinking)	Может быть уменьшен	Может быть уменьшен	Зависит от количества активных VLF

13

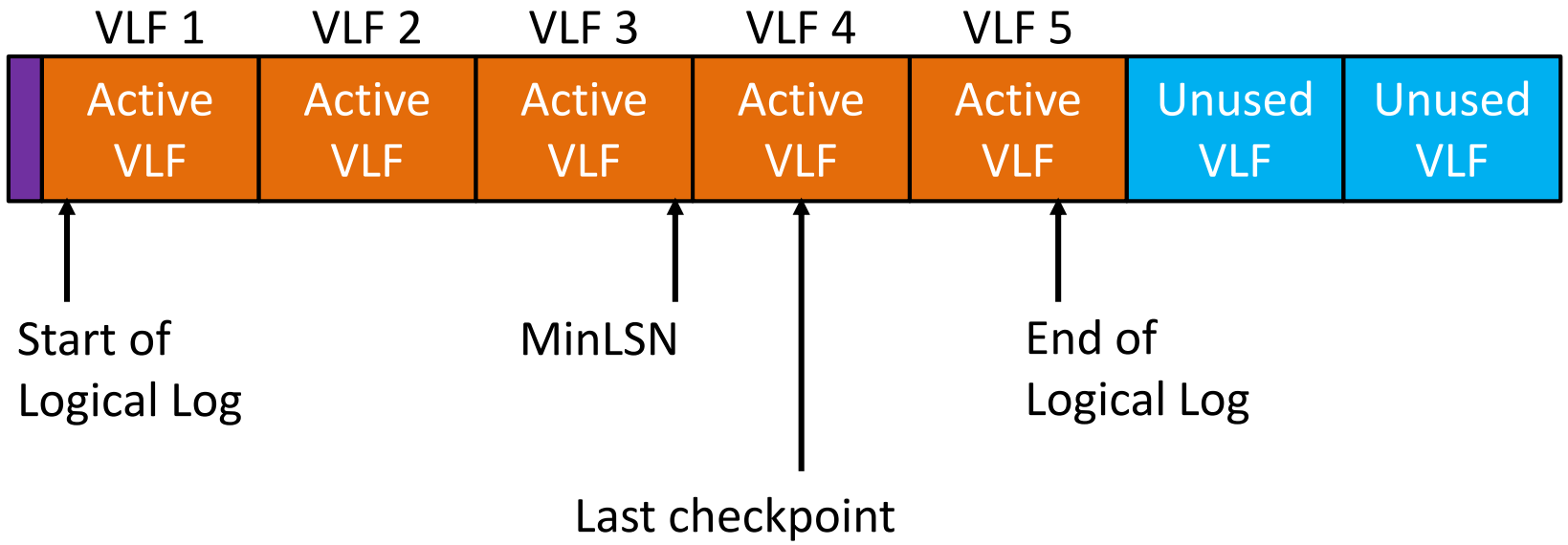
Расширение файла журнала (1)



- SQL сервер выделяет столько VLF, сколько нужно для отката самой длинной транзакции
- Новые VLF всегда заполняются нулями

14

Расширение файла журнала (2)



- SQL сервер выделяет столько VLF, сколько нужно для отката самой длинной транзакции
- Новые VLF всегда заполняются нулями

15

Алгоритм определения размера VLF

- Используется при создании файла лога во всех версиях SQL сервера
- Используется при расширении файла лога в старых версиях (до SQL сервер 2012 включительно)
- Зависит от размера, на который будет расширен файл журнала
- Размер каждого нового VLF примерно одинаков

Размер расширения файла журнала	Количество добавляемых VLF
До 1 Мбайт включительно	До 4 VLF, первые VLF по 248 Кбайт
Более 1 Мбайт – до 64 Мбайт включительно	4 VLF
Более 64 Мбайт – до 1 Гбайт включительно	8 VLF
Более 1 Гбайт	16 VLF

16

Новый алгоритм определения размера VLF

- Используется только при расширении файла лога в версиях, начиная с SQL сервера 2014
- Зависит от размера, на который будет расширен файл журнала
  - Если размер расширения файла меньше 1/8 текущего размера, будет создан 1 VLF
  - Если больше, то используется старый алгоритм

17

Параметры расширения файла журнала

- Файл журнала имеет начальный и максимальный размеры
- Максимальный размер может быть фиксированным или неограниченным*
- Файл журнала может расширяться вручную или автоматически
- Если происходит расширение файла:
  - Создаются новые VLF и заполняются нулями
  - Вызывает задержку в обработке транзакций

18

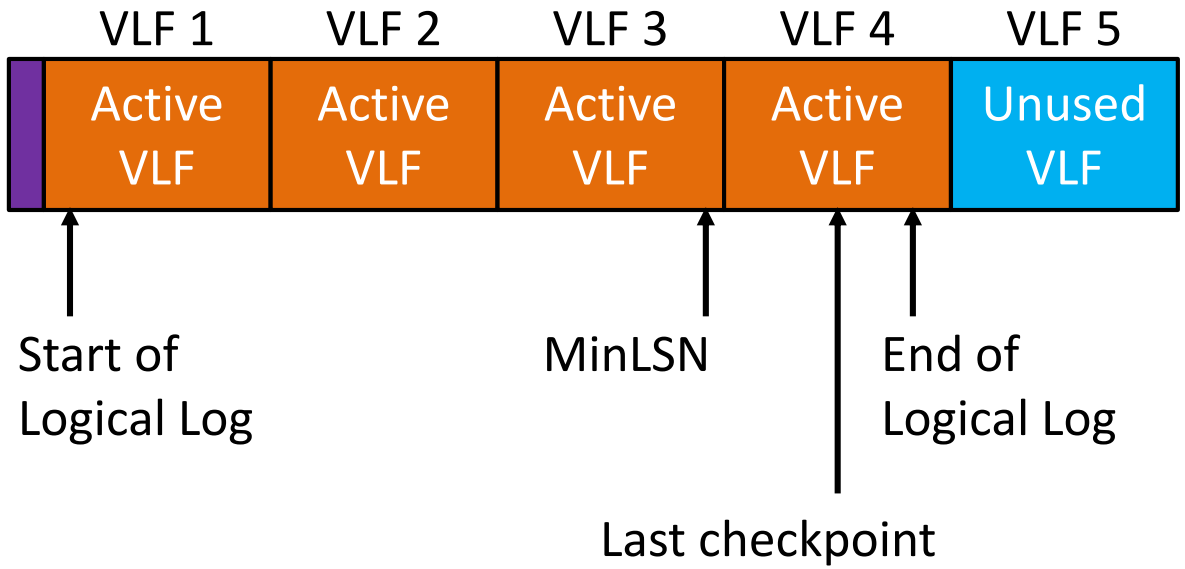
Неактивные записи журнала

Запись журнала становится неактивной в случае, если:

- Транзакция, к которой относится запись, подтверждена
- Страница с данными, которая была изменена этой транзакцией, была записана на диск в результате выполнения контрольной точки (checkpoint)
- Запись журнала не нужна для выполнения резервной копии БД или лога
- Запись журнала не нужна для любого механизма, который читает лог (зеркалирование, AlwaysOn Availability Group, транзакционная репликация, Change Data Capture)

19

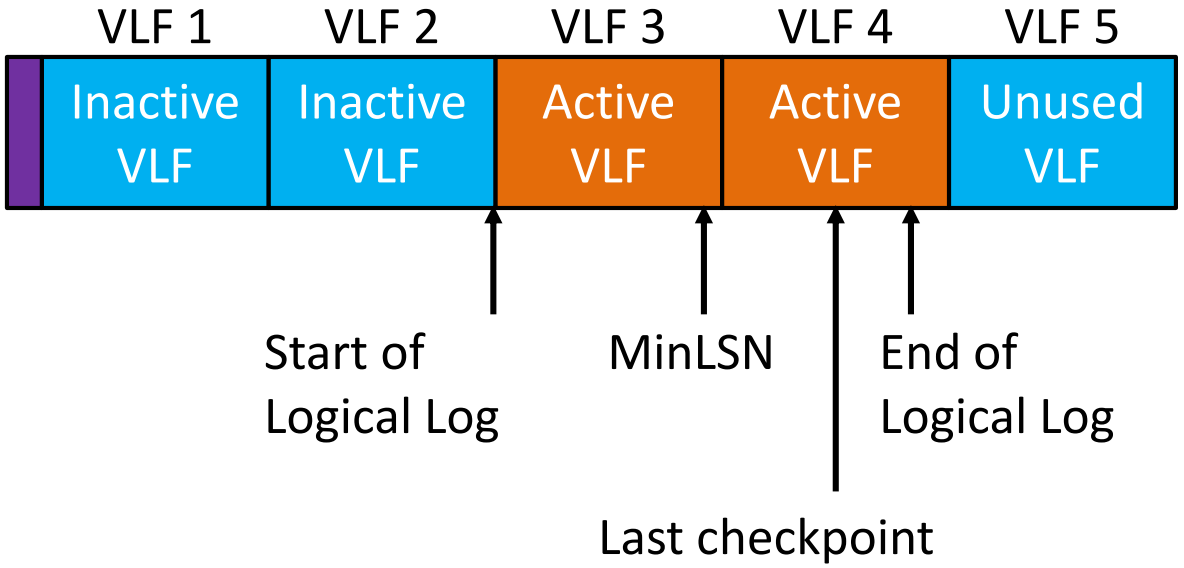
Усечение журнала (1)



- VLF усекается, если:
  - Не содержит активных записей лога
  - После контрольной точки в моделях восстановления simple/pseudo-full
  - После резервной копии лога в моделях восстановления full/bulk-logged

20

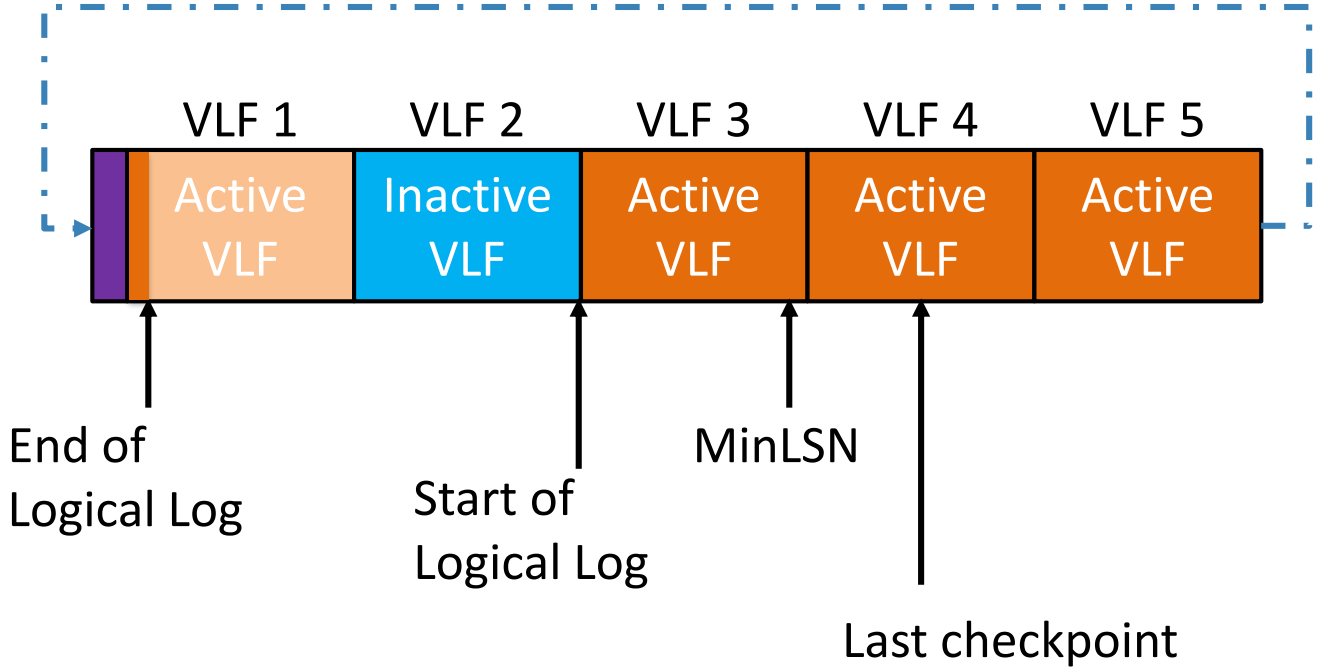
Усечение журнала (2)



- VLF помечен как усеченный
- VLF не перезаписывается нулями

21

Циклическая перезапись журнала



- Неактивные VLF могут быть перезаписаны
- Первый VLF обязательно должен быть неактивным

22

Сжатие файла

- Автоматическое
  - Опция базы данных AUTO_SHRINK
- Ручное
  - Выполняется командой DBCC SHRINKFILE
  - Сжимается только свободное место в конце файла (за счет неактивных VLF)
  - Можно сжать максимум до первых 2 VLF

23

Устранение проблем

- Переполнение журнала (ошибка 9002)
- Сжатие файла лога
- Фрагментация журнала транзакций

24

Чрезмерный рост журнала

Причины:

- Как узнать? SELECT log_reuse_wait_desc FROM sys.databases
- Описание в разделе «Factors That Can Delay Log Truncation» в статье https://msdn.microsoft.com/en-us/ms190925.aspx

Как исправить:

Причина	Возможное решение
LOG_BACKUP	Установить более частое резервное копирование лога
ACTIVE_BACKUP_OR_RESTORE	Изменить стратегию/расписание резервного копирования
ACTIVE_TRANSACTION	Убить «тяжелую» транзакцию
REPLICATION	Проверить транзакционную репликацию

25

Мониторинг размера журнала

- Счетчики Performance Monitor (раздел SQLServer:Databases)
  - Log File(s) Size (KB)
  - Log File(s) Used Size (KB)
  - Percent Log Used
  - Log Growths
- DBCC SQLPERF(LOGSPACE)
- sys.dm_db_log_space_usage (начиная с версии SQL Server 2012)

26

Ошибка 9002

Если файл лога не может быть расширен автоматически:

- Возникнет ошибка 9002
- Откат всех незавершенных транзакций
- Остановка работы SQL сервера

Как исправить:

- Посмотреть причину, затем выполнить соответствующее действие
- Расширить файл журнала (если возможно)
- Добавить дополнительный файл журнала

27

Демонстрация

- Переполнение журнала и ошибка 9002
- Удаление лишних файлов журнала

28

Сжатие журнала

Последовательность:

- Выполнить DBCC LOGINFO для оценки количества VLF и последнего активного VLF
- Сделать усечение лога
  - Резервная копия лога
  - Контрольная точка (checkpoint)
- Дождаться перехода активной части лога в начало файла и сделать усечение повторно
- Выполнить DBCC SHRINKFILE

29

Фрагментация VLF

- VLF добавляются в процессе расширения лога
  - Неправильные значения расширения лога могут привести к созданию большого числа маленьких VLF
- Усеченные VLF могут быть в любом месте файла лога
  - Приводит к фрагментации в логической последовательности VLF
- Вызывает проблемы в производительности журнала транзакций, резервном копировании/восстановлении, операциях чтения лога
- Если количество VLF составляет сотни или тысячи, стоит задуматься о дефрагментации VLF

30

PASS

Дефрагментация VLF

- Выполнить сжатие файла лога
- Повторить сжатие до достижения минимального размера
- Изменить размер журнала транзакций и/или параметры авторасширения
  - Размер VLF не следует делать больше 500 Мбайт
  - При ручном расширении увеличивать размер файла максимум на 8 Гбайт

31

Демонстрация

- Сжатие файла лога
- Дефрагментация VLF

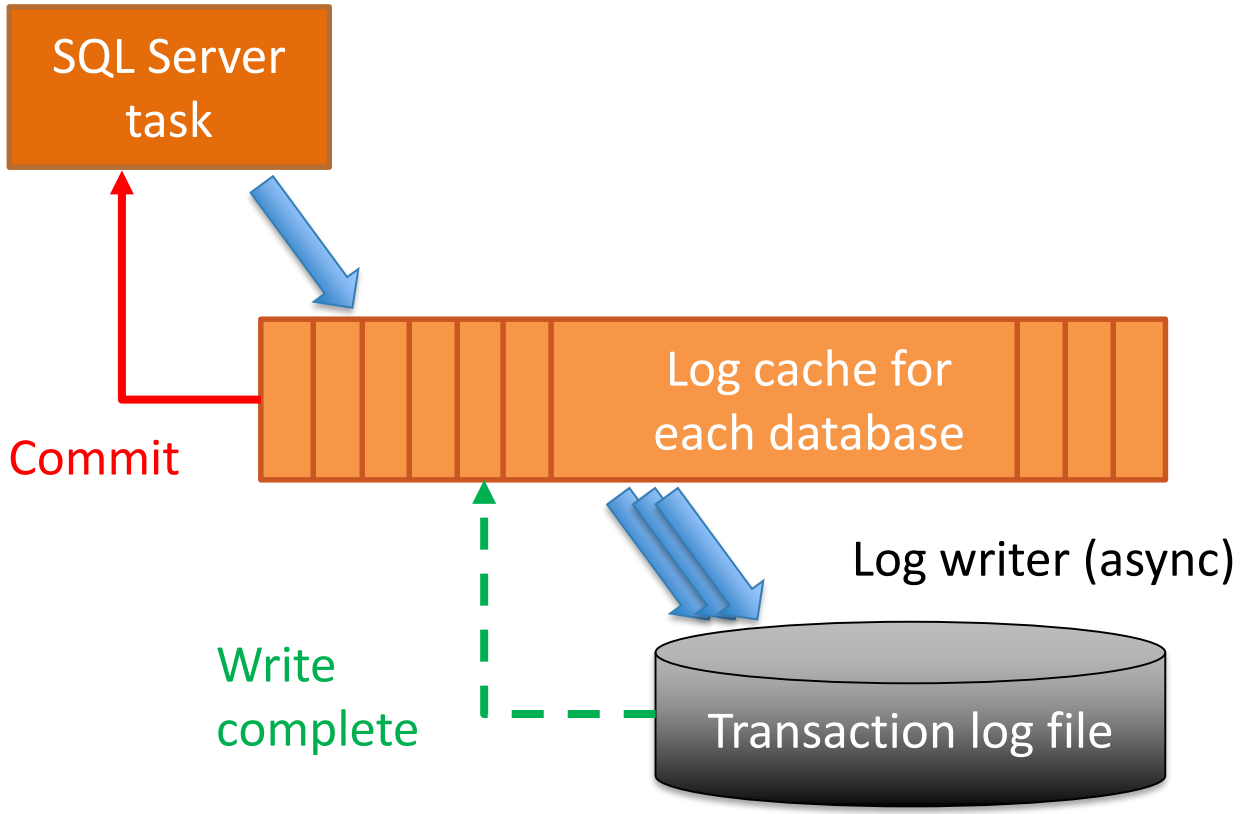
32

Отложенные транзакции (Delayed Durability)

- Версия SQL сервера – 2014 и выше
- Подтверждение транзакции (commit) происходит до записи на диск
- Задается на уровне базы данных
- Преимущества:
  - Сокращение ожиданий
  - Увеличение пропускной способности за счет большего размера блоков журнала
- Недостатки:
  - Риск потери данных, нарушение целостности в случае сбоя

33

Запись отложенных транзакций



34

Демонстрация

- Производительность отложенных транзакций

35

Рекомендации по администрированию

- Разместить журнал транзакций на высокоскоростном отдельном физическом диске/дисковой подсистеме
- Только один файл журнала
- Не использовать параметр авторасширения файла в процентах
- Следить за производительностью и ростом журнала
- Предотвращать переполнение журнала
- Управлять количеством VLF
- Подумайте о переходе на новую версию SQL сервера

36

Рекомендации по производительности

- Увеличить размер транзакций
- Удалить неиспользуемые индексы
- Сократить количество разбиений страниц за счет перестроения индексов с другим Fillfactor
- Оценить производительность/задержку решений HA&DR, которые используют журнал транзакций
- Подумайте об использовании возможностей SQL сервера 2014
  - Отложенные транзакции (Delayed Durability)
  - Оптимизация в памяти (In-Memory OLTP)

37

Вопросы?

Спасибо за внимание!