

Диагностика проблем в SQL Server - практический подход

Докладчик, плохо владеющий терминологией

- 11+ лет работы с Microsoft SQL Server
- Microsoft SQL Server MVP
- Microsoft Certified Master (SQL Server 2008)
- MCPD
 - Enterprise Application Developer



- Blog: <http://aboutsqlserver.com>
 - Сессия будет доступна для скачивания
- Email: dmitri@aboutsqlserver.com



О чем мы будем говорить?

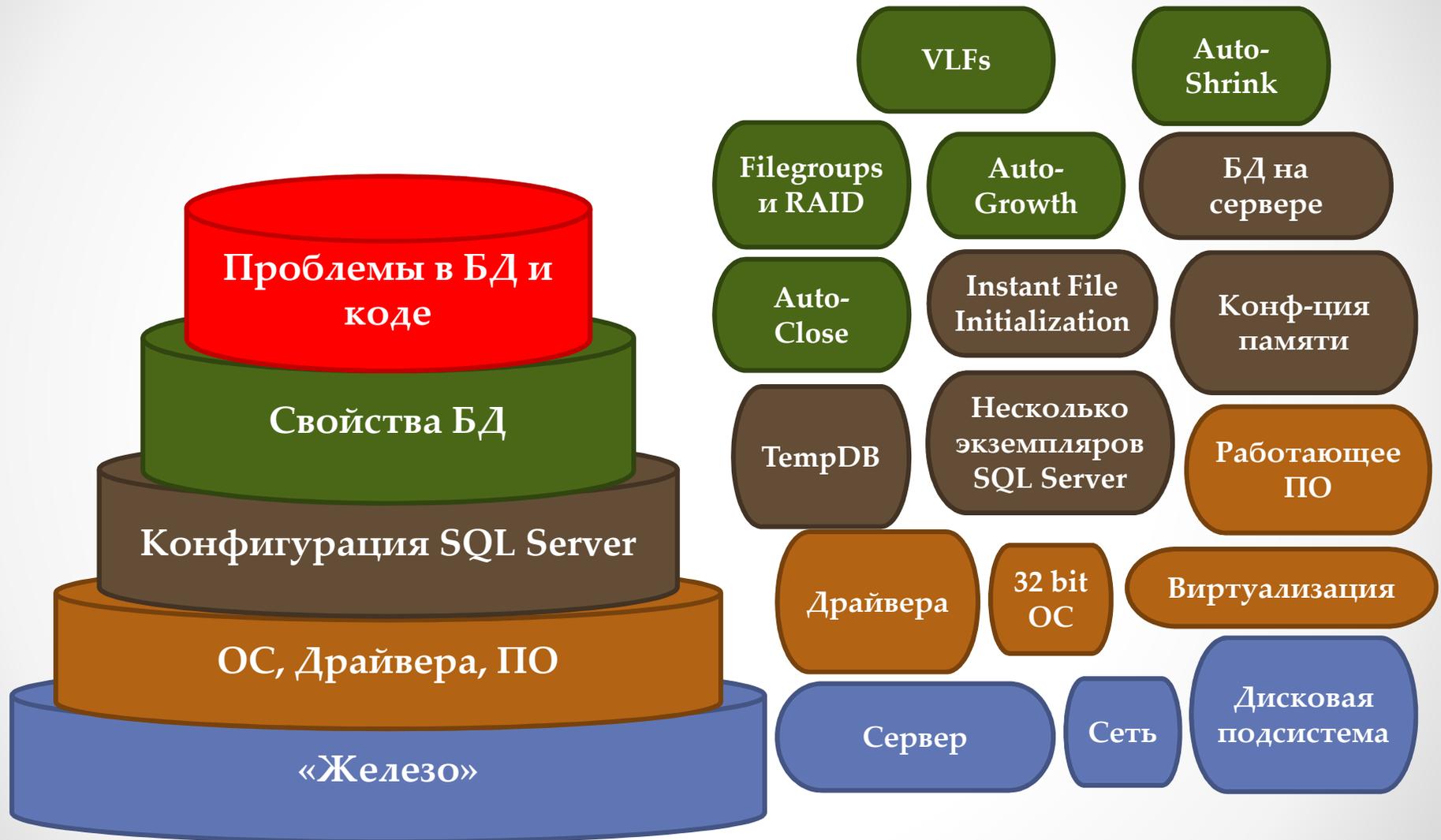
- **Мы поговорим о:**

- Модели выполнения SQL Server
- Методике диагностики проблем с помощью статистики ожиданий, системных счетчиков и DMV
- Как разные проблемы проявляют себя в системе

- **Цель:**

- Поделиться опытом
- Продемонстрировать как мы можем диагностировать проблемы в OLTP системах

Не забываем о «мелочах»



Полная Картина (1)

- Железо
 - Достаточно ли мощный сервер?
 - Дисковая подсистема
 - RAID
 - Пропускная способность (тестирование с помощью SQLIO/SQLIOSim)
 - Выровнен ли диск? Какой размер сектора?
 - Пропускная способность сети
- ОС
 - Все ли в порядке с драйверами
 - 32 битная ОС – конфигурация памяти (AWE, /3GB /UserVA)?
 - Min/Max server memory и “Lock Pages in Memory” ?
 - Какой софт работает на сервере?
 - Виртуальный ли сервер? Используется ли balloon драйвер? Что происходит с хостом?

Полная Картина (2)

- Конфигурация SQL Server
 - Несколько экземпляров?
 - Несколько БД?
 - Совпадает ли воркload (OLTP/DW)?
 - Audit/security/TDE?
 - TempDB
 - Используется ли быстрый диск?
 - Сколько файлов?
 - Какой начальный размер?
 - Правильно сконфигурирована ли память?
 - Включен Instant File Initialization?
- БД
 - Выключен ли Auto-shrink и Auto-close?
 - Достаточно ли свободного места в журнале транзакций? Сколько VLF?
 - Какие параметры автоматического роста журнала транзакций?
 - Как сконфигурированы файловые группы и файлы?

Базовая линия (Baseline)

- Для нас
 - Отдельные значения метрик не всегда информативны
 - “В системе происходит 25 сканирований таблиц в секунду..Это нормально”?
 - “Задержки дисковой подсистемы составляют 20мс. Плохо ли это?”
 - Наблюдение за базовой линией позволяет узнать о проблемах на ранней стадии
- Для клиентов и начальства 😊
 - Что выглядит более профессионально: «Мы уменьшили нагрузку на процессоры» или «% ожиданий ресурсов процессора уменьшился с 50% до 15%”.

SQLLOS

- Прослойка между SQL Server и Windows
- Отвечает за:
 - Управление рабочими процессами
 - Работу с дисковой подсистемой
 - Управление памятью и ресурсами

Модель выполнения в SQL Server

- SQLOS создает 1 диспетчер (scheduler) на **ЛОГИЧЕСКИЙ** процессор
- Рабочие потоки (Worker Threads) распределяются между диспетчерами
- Запрос распределяется между одним или несколькими рабочими потоками.
Распределение не меняется до окончания обработки запроса
- Состояния рабочих процессов:
 - Выполняется (Running) – выполняется в текущий момент
 - Заблокирован (Suspended) – ожидание ресурсов
 - Ожидание выполнения (Runnable) – ожидание процессора

1 диспетчер

Кассир =
диспетчер



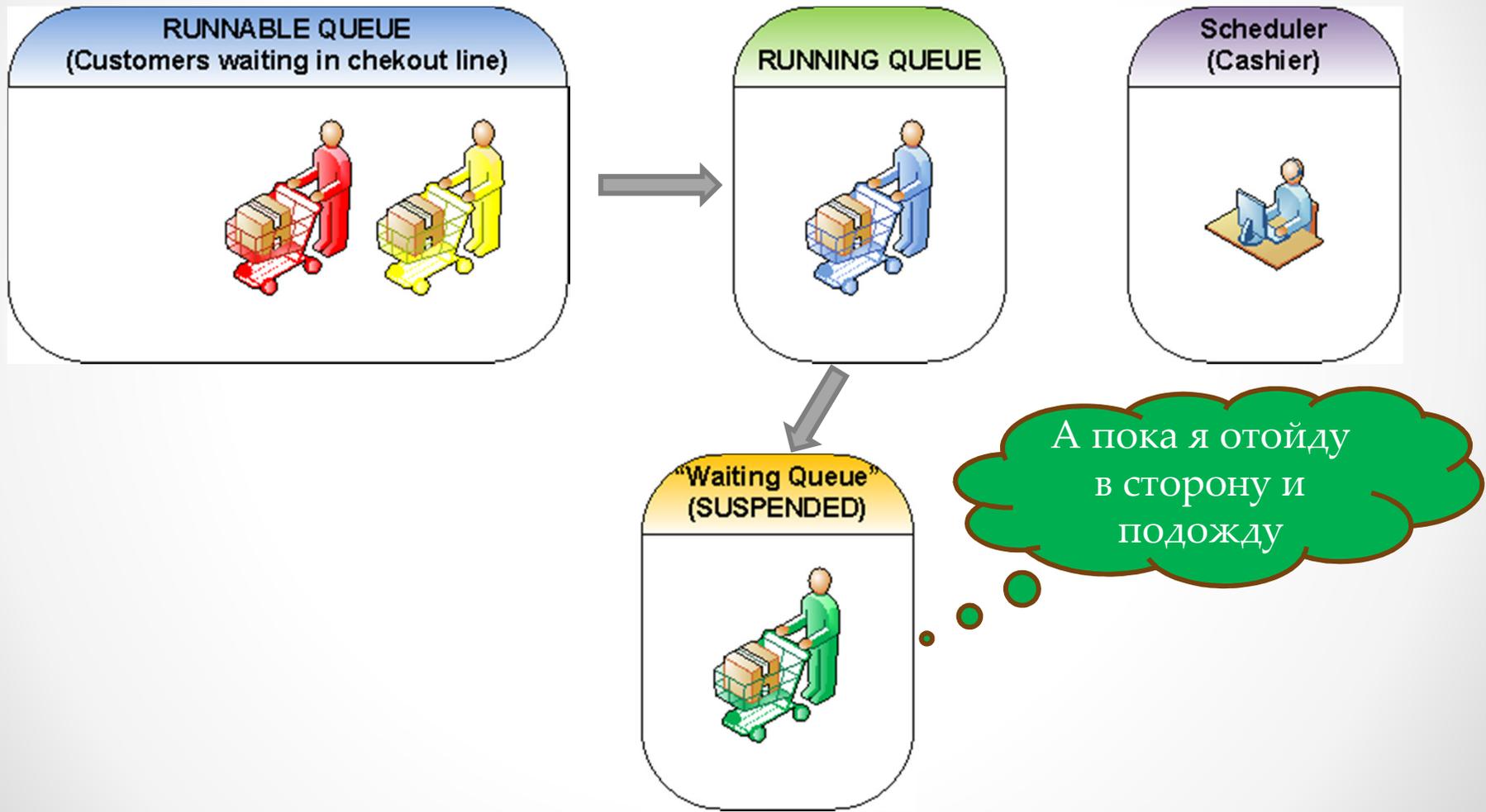
1 диспетчер

У меня на
продукте нет
бар-кода

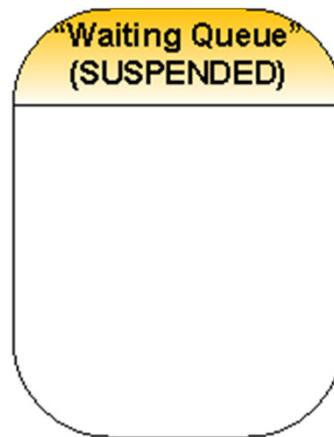
Я пошлю коллегу
узнать цену



1 диспетчер

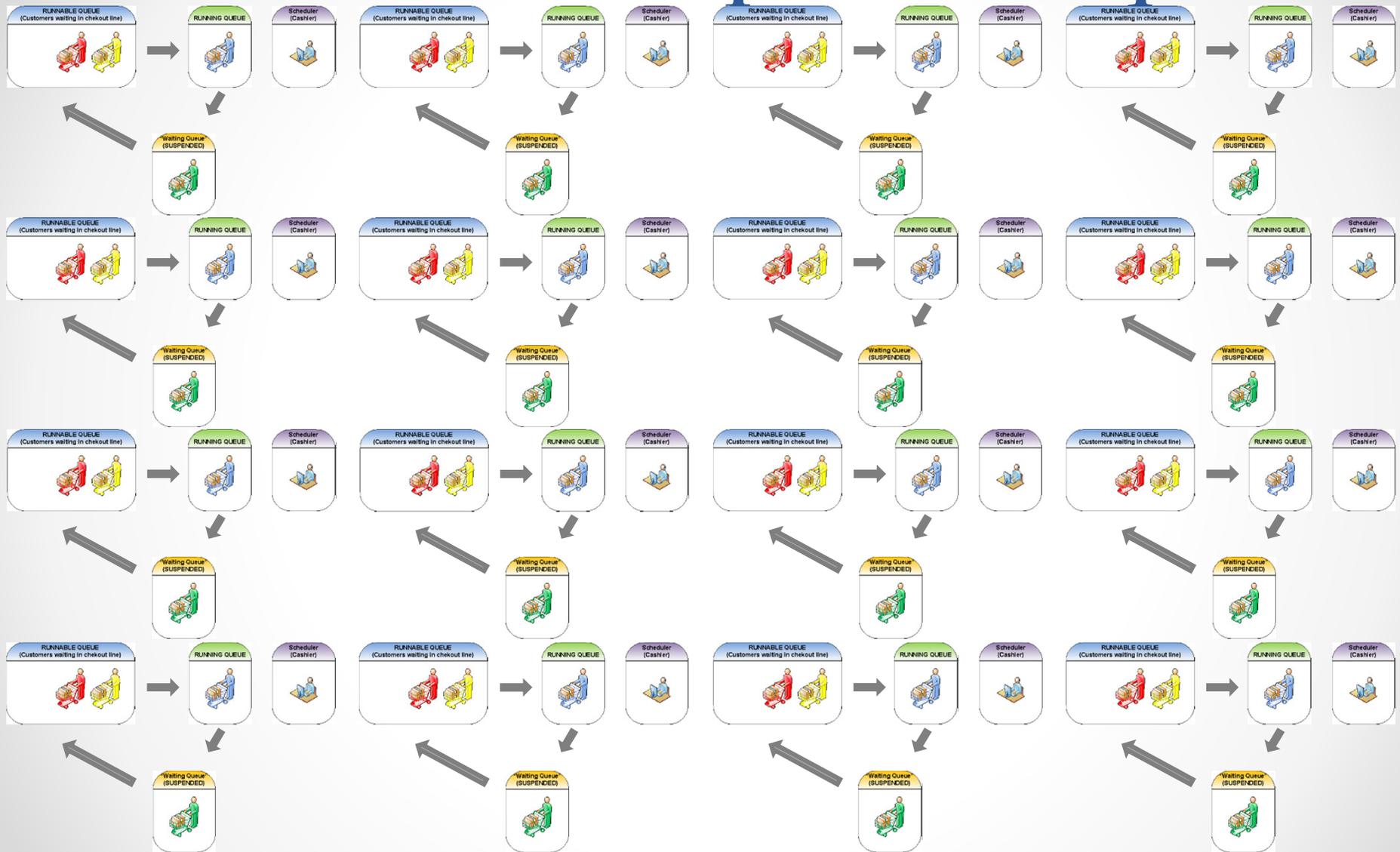


1 диспетчер

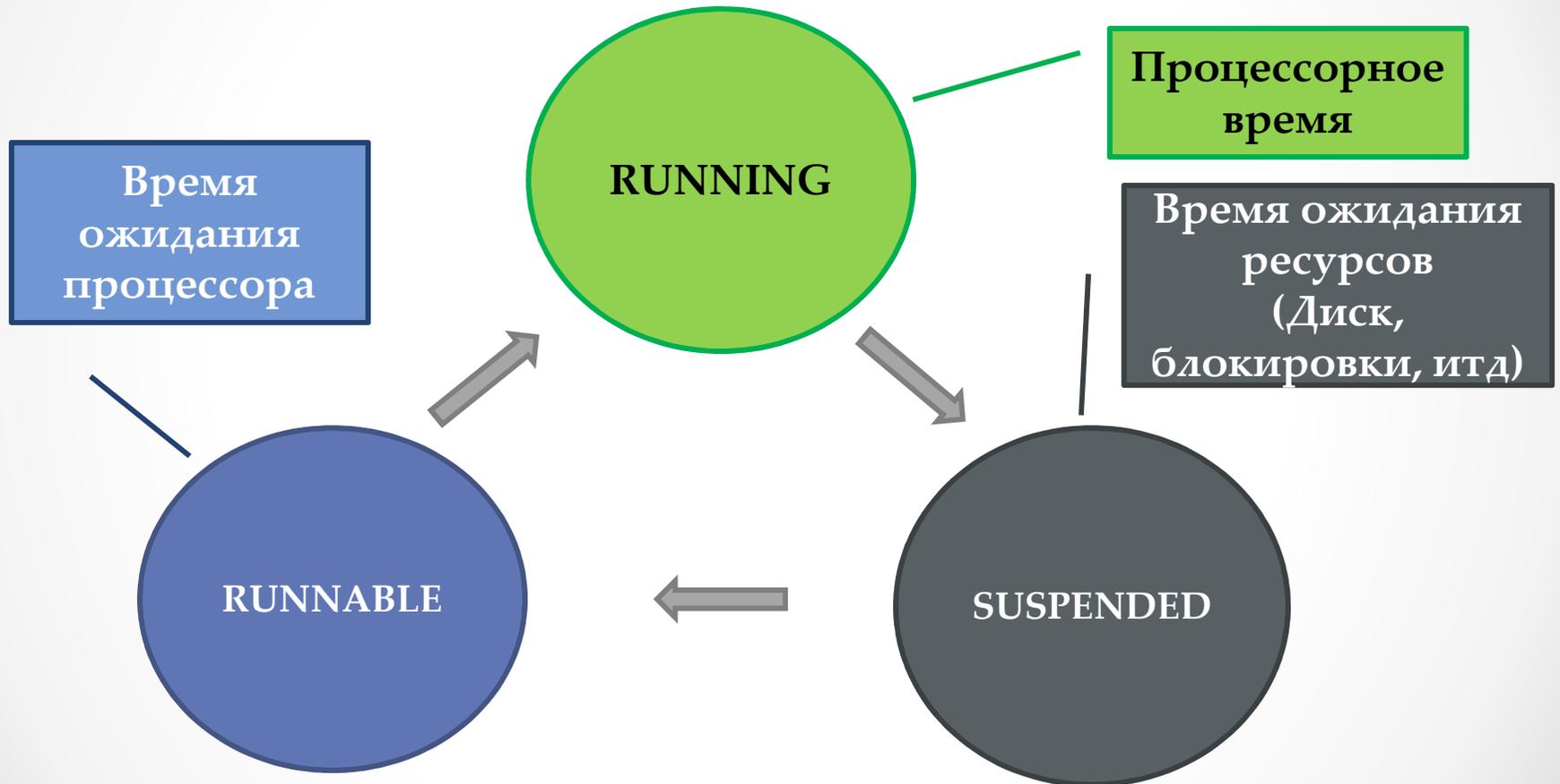


Узнав цену я
снова встану в
конец очереди
☹

Несколько процессоров?



Жизненный цикл запроса



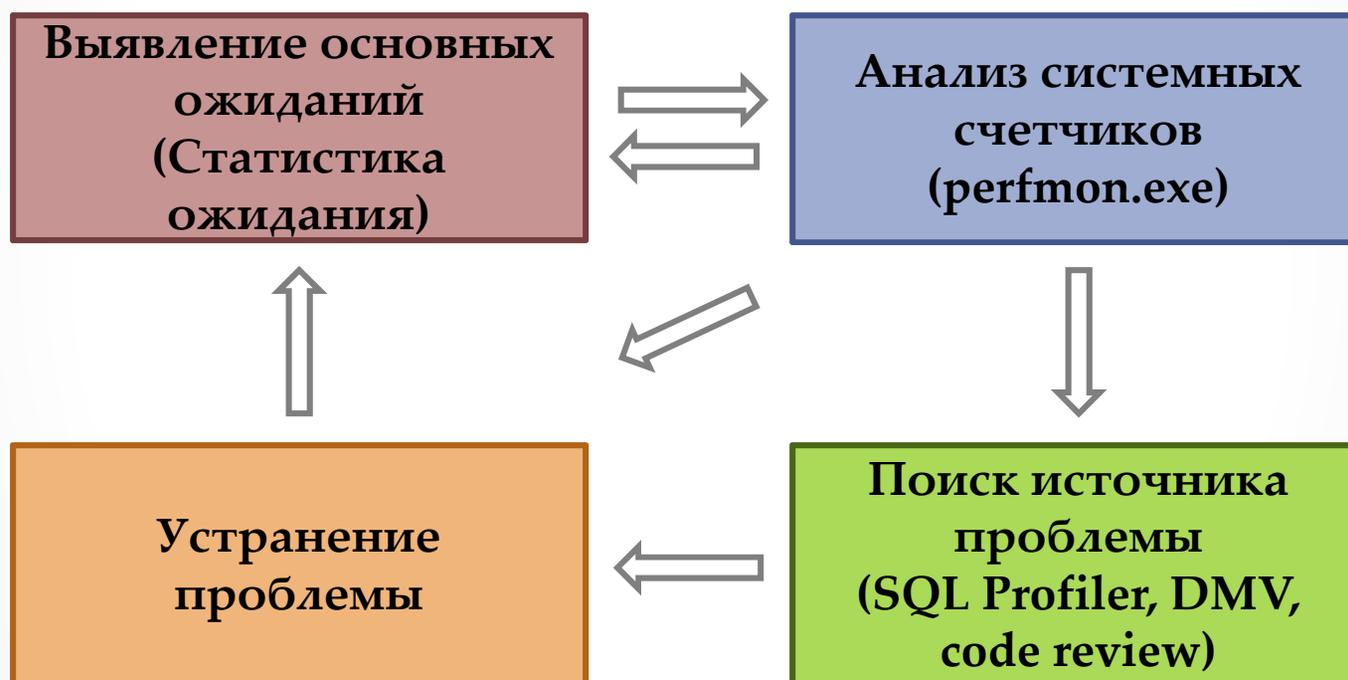
Статистика ожиданий

- Чего мы ждем?

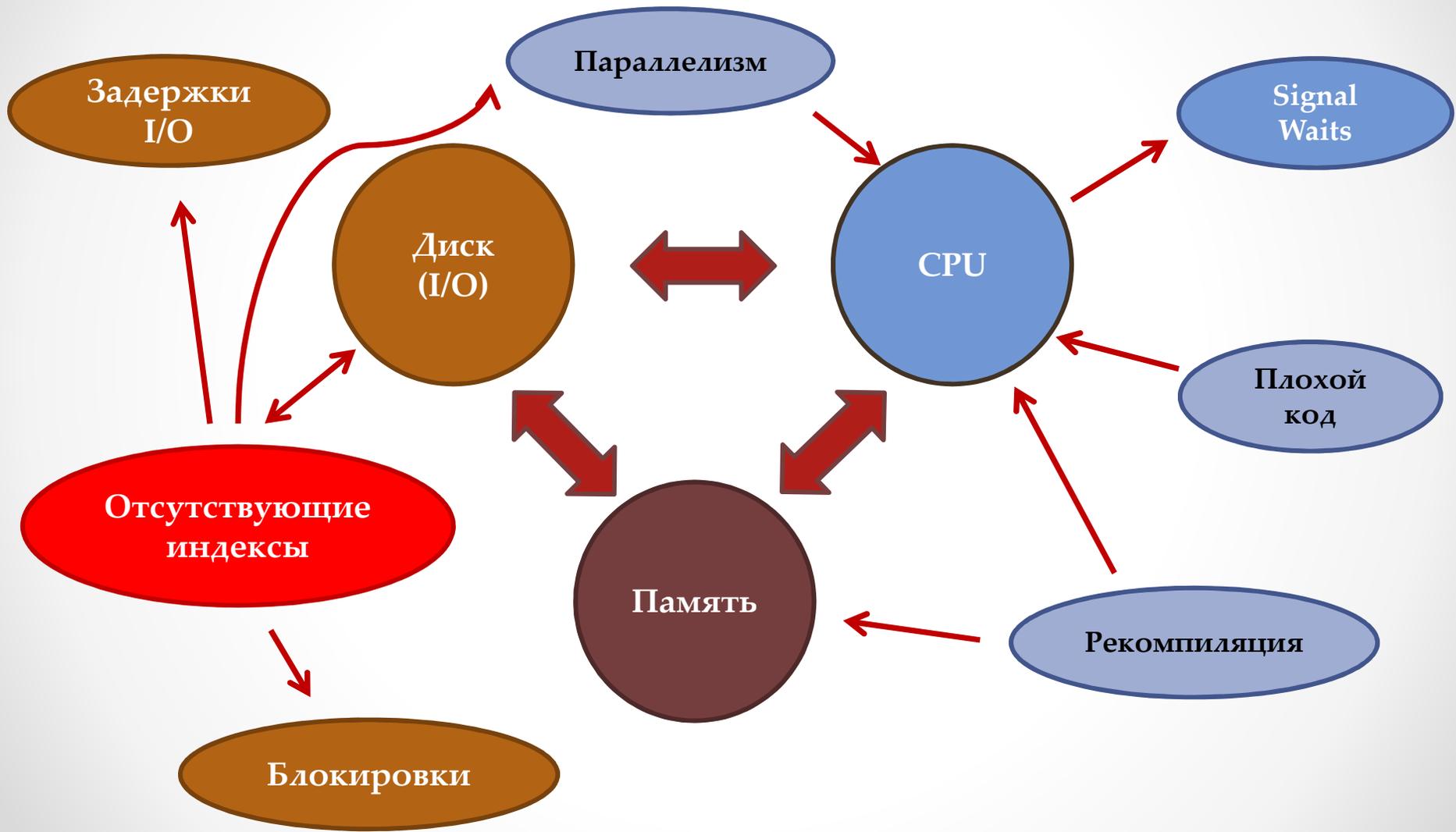
```
SELECT
    wait_type, wait_time_ms,
    convert(decimal(7,4), 100.0 * wait_time_ms / SUM(wait_time_ms) OVER()) AS [Percent]
from
    sys.dm_os_wait_stats
where
    wait_type NOT IN ('CLR_SEMAPHORE', 'LAZYWRITER_
    , 'SLEEP_SYSTEMTASK', 'SQLTRACE_BUFFER_FLUSH', 'W
    , 'REQUEST_FOR_DEADLOCK_SEARCH', 'XE_TIMER_EVENT
    'CLR_MANUAL_EVENT', 'CLR_AUTO_EVENT', 'DISPATCHER
    , 'XE_DISPATCHER_WAIT', 'XE_DISPATCHER_JOIN', '
    , 'OLEDB', 'MSQL_DQ' )
    and wait_type not like 'Broker%'
order by
    [Percent] Desc
```

	wait_type	wait_time_ms	Percent
1	BACKUPBUFFER	11311975189	20.6899
2	BACKUIO	11153312594	20.3997
3	PAGEIOLATCH_EX	4620890485	8.4517
4	WRITELOG	3983896698	7.2866 JE'
5	PAGEIOLATCH_SH	3719440813	6.8029
6	CXPACKET	3630197534	6.6397
7	OLEDB	3464571854	6.3368
8	MSQL_XP	2591424522	4.7398
9	ASYNC_IO_COMPLETION	2028625085	3.7104
10	SOS_SCHEDULER_YIELD	1870324254	3.4209
11	RESOURCE_SEMAPHORE	1314585339	2.4044
12	MSQL_DQ	1072516276	1.9617
13	LCK_M_U	816581103	1.4935
14	ASYNC_NETWORK_IO	598159386	1.0940
15	PAGFIATCH SH	351490272	0.6429

Вечный цикл

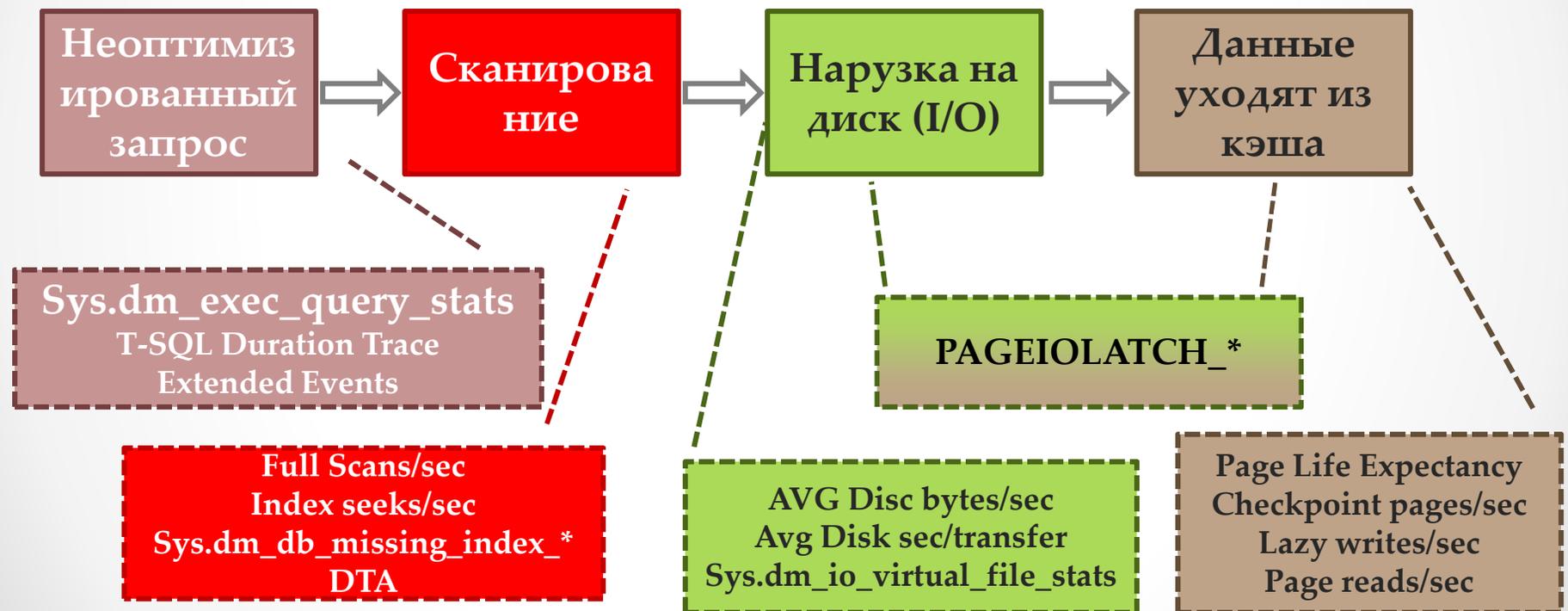


Все взаимосвязано



Проблемы с памятью и дисковой подсистемой

- Почти всегда связаны с неоптимизированными запросами



Проблемы с памятью и ДИСКОМ

Type	Name	Description
Типы ожиданий:	PAGEIOLATCH_*	Чтение данных с диска в кэш (Physical IO)
	IO_COMPLETION, ASYNC_IO_COMPLETION	Чтение с диска обычно не связанное с данными
	WRITELOG, LOGMRG	Дисковые операции с журналом транзакций
Системные счетчики:	Buffer cache hit ratio	Как часто данные находятся в кэше. Не использовать
	(Avg) Disk Queue Length	Размер очереди на операции с диском
	Page life expectancy	Сколько времени страница находится в кэше. Анализируем тренд. Минимальное значение: $> (DB_CACHE_SIZE / 4GB) * 300 \text{ sec.}$
	Checkpoint pages/sec Lazy writers/sec	Как часто данные записываются на диск Проблемы: большие значения + низкий PLE
	Page reads/sec	Количество страниц прочитанных в секунду
	Avg Disk Bytes/* Avg Disk sec / Transfer	Счетчики производительности диска

Проблемы с памятью и ДИСКОМ

Type	Name	Description
Тип ожиданий:	RESOURCE_SEMAPHORE	Информация о выделении памяти запросам Должны быть минимальны в OLTP Нормальны для Data Warehouse
Системные счетчики:	Memory Grant Pending	
	Memory Grant Outstanding	
DMV:	<u>sys.dm_exec_query_stats</u>	Статистика выполнения запросов
	<u>sys.dm_io_virtual_file_stats</u>	Статистика по производительности дисковой подсистемы на уровне файлов БД. Io_stall – время ожидания дисковой подсистемы
	sys.dm_os_memory_clerks DNCC MEMORYSTATUS	Что использует память

Sys.dm_exec_query_stats

```
SELECT TOP 250
    SUBSTRING(qt.TEXT, (qs.statement_start_offset/2)+1,
        ((
            CASE qs.statement_end_offset
            WHEN -1 THEN DATALENGTH(qt.TEXT)
```

	SQL	Exec ...	Avg IO	query_plan	Total Reads	Total Writes	Total CPU
1	select Subj, cast(R...	1	6816382	<ShowPlanXM...	6816296	86	24297389
2	select UID, DOCTY...	26455	4143503	<ShowPlanXM...	109616393555	0	154369131409
3	DELETE TOP (@d...	1	4096631	<ShowPlanXM...	4096468	163	26538518
4	insert into #tmpRep...	62	3690210	NULL	228750206	42859	3351099613
5	update #tmpReportl...	62	3139967	NULL	194677952	7	2406888686
6	insert into #tmpRep...	58	2516483	NULL	145905711	50341	1761652781
7	select D.*, O.CATE...	16	1848720	<ShowPlanXM...	29579527	0	64629691
8	update #tmpReport...	13	1520333	<ShowPlanXM...	19764334	5	194722131
9	select D.*, I.Catego...	36	1511917	<ShowPlanXM...	54429042	0	114735561
10	update #tmpReport...	26	1459946	<ShowPlanXM...	37958482	138	447010567
11	update #tmpReport...	12	1426777	<ShowPlanXM...	17121325	4	164099386
12	insert into #tmpRep...	53	1079374	NULL	57198359	8467	865721533

```
ORDER BY
    [Avg IO] desc
option (recompile)
```

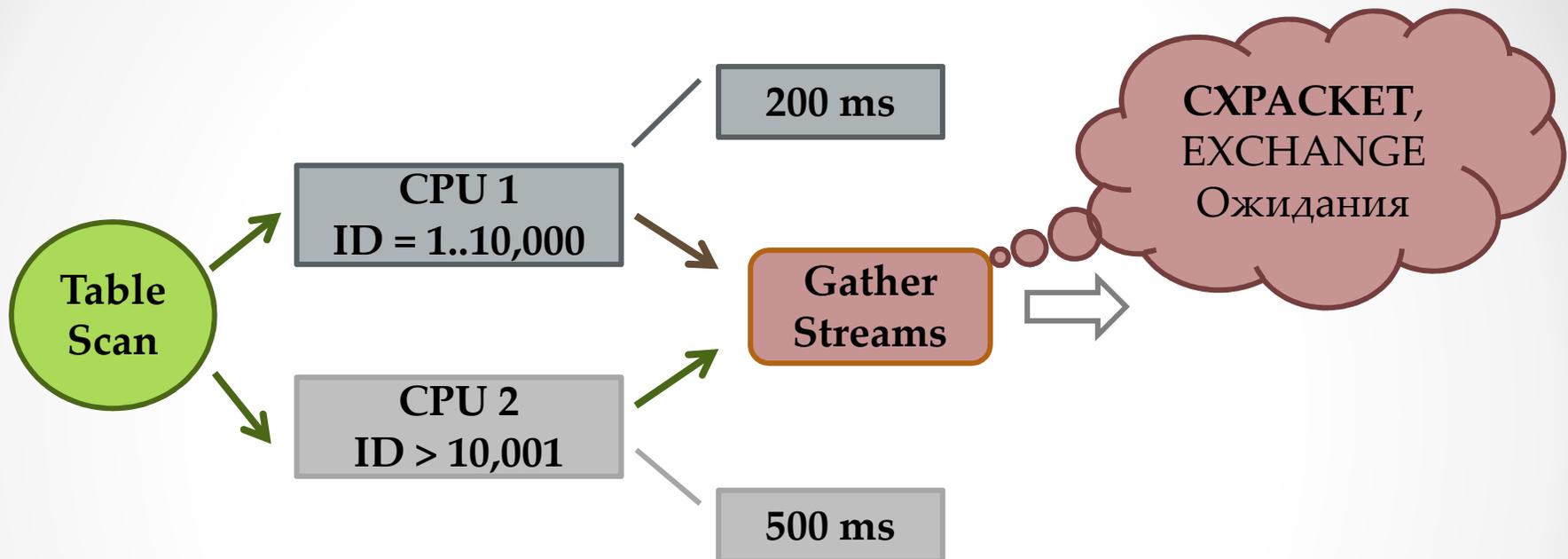
!! Работает только с закэшированными планами !!

Диагностика проблем связанных с дисковой подсистемой



Демонстрация

Параллелизм



- Параллелизм обычно не нужен в OLTP
- Параллелизм всегда присутствует в Data Warehouse
- MaxDOP должен быть \leq # CPUs в NUMA ноде
- Рекомендуется увеличить "Cost Threshold for Parallelism" вместо уменьшения MaxDOP в OLTP

Диагностика проблем связанных с параллелизмом



Демонстрация

Проблемы с CPU

Type	Name	Description
Типы ожиданий:	SOS_SCHEDULER_YIELD	Ожидание процессора
	CMEMTHREAD	Возможно Ad-hoc sql
DMV:	Sys.dm_os_wait_stats	Signal_wait_time_ms > 25%
	sys.dm_os_memory_clerks	CACHESTORE_SQLCP: Ad-Нос планы выполнения
Системные счетчики:	Batch Requests/sec	Количество запросов в секунду
	SQL Compilations/sec	Количество компиляций
	SQL Re-Compilations/sec	Количество перекомпиляций

- Потенциальные проблемы:
 - Ad-Нос SQL / Динамический SQL / Перекомпиляция
 - Плохой код T-SQL
 - Неоптимизированные запросы
- OLTP системы:
 - Initial Compilations = Sql Compilations/sec – SQL Re-Compilations/sec
 - Переиспользование плана = $(\text{Batch requests/sec} - \text{Initial Compilations}) / \text{Batch request/secs} \geq 90\%$

Диагностика проблем связанных с рекомпиляцией

• • •

Демонстрация

Скалярные функции

...

Демонстрация

Async_Network_IO

- Клиент «читает» данные недостаточно быстро
- Потенциально:
 - Проблемы с сетью.
 - Проблемы с клиентским кодом
 - **Зачитывайте все данные перед обработкой!**

Async_Network_IO

• • •

Демонстрация

Проблемы, связанные с блокировками

Type	Name	Description
Типы ожиданий:	LCK_M_*	Ожидания на получение блокировок
DMV:	Sys.dm_tran_locks	Текущие блокировки
Трейсы & Extended Events	Blocked Process Report	
	Deadlock graph	
Системные счетчики:	<Instance>\Locks	Locks/Timeouts/Deadlocks statistics

Диагностика проблем связанных с блокировками



Демонстрация

Эскалация блокировок

- SQL Server переносит блокировки на уровень таблиц/секций
 - Первоначально: ~5,000 блокировок на объект
 - Если не получилось – повторяет попытку через ~1,250 блокировок
- Проблема: пакетная операция создает эскалацию блокировок. Все остальные сессии не имеют доступа к объекту
- Диагностика
 - Высокий процент ожиданий на «бюлокировки по намерению» (LCK_M_I*)
 - SQL Profiler Locks: Escalation
- Решение
 - Trace flag 1211 (instance level) – Не рекомендуется, но иногда необходим
 - SQL Server 2008+: *alter table .. set lock_escalation*
 - Оптимистические уровни изоляции
 - Row version model – запись не блокирует чтение

Эскалация блокировок



Демонстрация

Реальная история

	wait_type	wait_time_...	Percent
1	CXPACKET	47237677	37.0492
2	LCK_M_IS	17641793	13.8367
3	PAGELATCH_UP	10757870	8.4375
4	LCK_M_SCH_S	10103857	7.9246
5	ASYNC_NETWORK_IO	9715441	7.6200
6	SOS_SCHEDULER_YIELD	8970275	7.0355
7	LCK_M_SCH_M	5748216	4.5084
8	OLEDB	3335574	2.6161
9	LCK_M_IX	3000305	2.3532
10	LATCH_EX	2621557	2.0561
11	ASYNC_IO_COMPLETION	1613775	1.2657
12	BACKUIO	1443624	1.1323
13	IO_COMPLETION	1115441	0.8749
14	BACKUPBUFFER	902306	0.7077
15	WRITELOG	882498	0.6922

- **Симптомы:**

- Высокий % блокировок на уровне схемы
- Высокий % ожиданий, связанных с параллелизмом
- Практически отсутствуют ожидания, связанные с диском

- **Шаг 1:**

- Диагностика блокировок на уровне схемы

- **Результат:**

- Полнотекстовый индекс обновляется после каждого изменения документов

Реальная история

	wait_type	wait_time_...	Percent
1	CXPACKET	6059039	44.1425
2	ASYNC_IO_COMPLETION	1747127	12.7285
3	BACKUIO	1483546	10.8082
4	BACKUPBUFFER	866660	6.3140
5	ASYNC_NETWORK_IO	573897	4.1811
6	SOS_SCHEDULER_YIELD	471540	3.4354
7	BACKUPTHREAD	436083	3.1770
8	LATCH_EX	417119	3.0389
9	IO_COMPLETION	331552	2.4155
10	LCK_M_S	299947	2.1852
11	WRITELOG	258726	1.8849
12	LCK_M_U	151601	1.1045
13	PAGEIOLATCH_EX	150622	1.0973

Симптомы:

- Высокий % ожиданий, связанных с параллелизмом
- Высокий % процент Signal Waits
- Практически отсутствуют ожидания, связанные с диском
- Нет признаков недостатка памяти

Проблемы:

- Большое количество неоптимизированных запросов
- Избыточное использование UDF
- База практически полностью закеширована
 - Нет необходимости в физическом I/O

Суммируя

- Проблемы с I/O
 - Оптимизируем запросы
- Проблемы с I/O и памятью
 - Оптимизируем запросы
- Проблемы с памятью без I/O
 - Определяем, что использует память
 - Открываем Google 😊
- Параллелизм в OLTP системах
 - Оптимизируем запросы
 - Увеличиваем “Cost Threshold for Parallelism”
- Блокировки
 - Оптимизируем проблематичные запросы
 - Решаем проблемы с эскалацией блокировок
 - Как временное решение осторожно переключаемся в READ COMMITTED SNAPSHOT
 - Проверяем клиентский код

Мы обсудили

- Модель выполнения SQL Server
- Методику диагностики проблем с помощью статистики ожиданий, системных счетчиков и DMV
- Увидели как разные проблемы проявляют себя в системе

Вопросы

- Спасибо за внимание!
- Сессия будет доступна для скачивания
 - <http://aboutsqlserver.com/presentations>

- Email: dmitri@aboutsqlserver.com